



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA STROJNÍHO INŽENÝRSTVÍ

FACULTY OF MECHANICAL ENGINEERING

ÚSTAV KONSTRUOVÁNÍ

INSTITUTE OF MACHINE AND INDUSTRIAL DESIGN

**REKONSTRUKCE 3D SCÉNY POMOCÍ PASIVNÍHO
STEREOPÁRU KAMER**

3D SCENE RECONSTRUCTION USING PASSIVE CAMERA STEREO-PAIR

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Leoš Fronc

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Jakub Hurník

BRNO 2019

Zadání bakalářské práce

Ústav: Ústav konstruování
Student: **Leoš Fronc**
Studijní program: Strojírenství
Studijní obor: Základy strojního inženýrství
Vedoucí práce: **Ing. Jakub Hurník**
Akademický rok: 2018/19

Ředitel ústavu Vám v souladu se zákonem č.111/1998 o vysokých školách a se Studijním a zkušebním řádem VUT v Brně určuje následující téma bakalářské práce:

Rekonstrukce 3D scény pomocí pasivního stereopáru kamer

Stručná charakteristika problematiky úkolu:

K zachycení 3D scény stačí za určitých podmínek pouze stereopár kalibrovaných kamer. Extrakce 3D informací z páru 2D snímků je důležitým procesem v počítačovém vidění. Metody umožňující zmíněnou rekonstrukci jsou se zvyšujícím se výpočetním výkonem a pokrokem ve výzkumu neuronových sítí stále zdokonalovány.

Typ práce: rešeršně syntetická

Cíle bakalářské práce:

Hlavním cílem je zpracování kritického přehledu současného stavu poznání v oblasti zpracování snímků z pasivního stereopáru kamer pro 3D rekonstrukci objektu.

Dílčí cíle bakalářské práce:

- zpracovat přehled metod rekonstrukce 3D scény z páru fotografií,
- porovnat metody z hlediska jejich klíčových vlastností,
- vytvořit jednoduchý praktický příklad využití některé z těchto metod.

Požadované výstupy: průvodní zpráva.

Rozsah práce: cca 27 000 znaků (15 – 20 stran textu bez obrázků).

Struktura práce a šablona průvodní zprávy jsou závazné:

http://dokumenty.uk.fme.vutbr.cz/BP_DP/Zasady_VSKP_2019.pdf

Seznam doporučené literatury:

SOLOMON, Chris, Toby BRECKON. Fundamentals of digital image processing: A practical approach with examples in Matlab. Hoboken, NJ: Wiley-Blackwell, 2011. ISBN 978-047-0844-731.

LUHMANN, Thomas. Close range photogrammetry: principles, techniques and applications. Dunbeath: Whittles Publishing, 2011. ISBN 978-184995-057-2.

Termín odevzdání bakalářské práce je stanoven časovým plánem akademického roku 2018/19

V Brně, dne

L. S.

prof. Ing. Martin Hartl, Ph.D.
ředitel ústavu

doc. Ing. Jaroslav Katolický, Ph.D.
děkan fakulty

ABSTRAKT

V bakalářské práci je vytvořen přehled současných metod pro 3D rekonstrukci scény pomocí pasivního stereopáru kamer. V práci jsou popsány a porovnány dva hlavní přístupy - triangulace významných bodů a výpočet disparitní mapy. Dále jsou vysvětleny jednotlivé metody v rámci těchto přístupů a ty jsou srovnány z hlediska jejich klíčových vlastností. Na základě srovnání jsou obecně vyvozeny závěry o volbě vhodných přístupů a metod pro konkrétní aplikace. Závěrem je vytvořen jednoduchý příklad, který ilustruje použití těchto metod v praxi.

KLÍČOVÁ SLOVA

3D rekonstrukce, stereopár kamer, triangulace významných bodů, disparitní mapa

ABSTRACT

This bachelor thesis presents an overview of current methods for the 3D reconstruction of the scene using a passive stereo camera. Two main approaches are described and compared in this thesis – the feature points triangulation and the disparity map. The following methods are described below and compared with their qualities. On the basis of this comparison, conclusions are drawn about the choice of suitable approaches and methods for specific applications. Finally, a simple example is created which illustrates the use of these methods in practice.

KEYWORDS

3D reconstruction, stereo camera, feature points triangulation, disparity map

BIBLIOGRAFICKÁ CITACE

FRONC, Leoš. *Rekonstrukce 3D scény pomocí pasivního stereopáru kamer* [online]. Brno, 2019 [cit. 2019-05-22]. Dostupné z: <https://www.vutbr.cz/studenti/zav-prace/detail/116640>. Bakalářská práce. Vysoké učení technické v Brně, Fakulta strojního inženýrství, Ústav konstruování. Vedoucí práce Jakub Hurník.

PODĚKOVÁNÍ

Na tomto místě bych rád poděkoval vedoucímu své bakalářské práce, panu Ing. Jakubu Hurníkovi za odborné rady, trpělivost a čas, který mi věnoval. Také bych rád poděkoval mé rodině za podporu během celého studia.

PROHLÁŠENÍ AUTORA O PŮVODNOSTI PRÁCE

Prohlašuji, že bakalářskou práci jsem vypracoval samostatně, pod odborným vedením Ing. Jakuba Hurníka. Současně prohlašuji, že všechny zdroje obrazových a textových informací, ze kterých jsem čerpal, jsou řádně citovány v seznamu použitých zdrojů.

V Brně dne.....

.....

Leoš Fronc

OBSAH

1	ÚVOD	13
2	ANALÝZA PROBLÉMU A CÍL PRÁCE	14
2.1	Analýza problému	14
2.2	Cíl práce	14
3	PŘEHLED SOUČASNÉHO STAVU POZNÁNÍ	15
3.1	3D rekonstrukce scény pomocí triangulace významných bodů	18
3.1.1	Detektory významných bodů	18
3.1.2	CNN detektory významných bodů	25
3.1.3	Srovnání detektorů	27
3.1.4	Deskriptory významných bodů	29
3.1.5	CNN deskriptory	34
3.1.6	Srovnání deskriptorů	35
3.1.7	Princip hledání korespondencí a triangulace	37
3.2	Disparitní mapa	39
3.2.1	Rektifikace	39
3.2.2	Výpočet míry podobnosti	40
3.2.3	Finální úprava	40
4	DISKUZE	41
4.1	Praktický příklad	42
5	ZÁVĚR	47
6	SEZNAM POUŽITÝCH ZDROJŮ	48
7	SEZNAM POUŽITÝCH ZKRATEK	51
8	SEZNAM OBRÁZKŮ	52
9	SEZNAM TABULEK	54

1 ÚVOD

Digitální 3D rekonstrukce scény z obrazových dat lze využít v širokém spektru aplikací počítačového vidění od automatické navigace robotů a automobilů, přes užití v průmyslu, až po aplikace v medicínské oblasti. Stále se zvyšující výpočetní výkon umožňuje širší implementaci počítačového vidění v těchto aplikacích, což motivuje vývoj v tomto oboru. Tím se také zvyšuje přesnost těchto metod a 3D rekonstrukci je možné použít například i pro oblasti přesného měření, kdy nejsme schopni objekty měřit konvenčními metodami nebo by tato možnost byla příliš nákladná či zdlouhavá. Nasazení počítačového vidění v průmyslu může vést ke snížení nákladů na výrobu a zlepšení efektivity výroby v důsledku automatizování zdlouhavých procesů, kterým je např. kontrola kvality.

K zachycení 3D scény stačí za určitých podmínek pouze stereopár kalibrovaných kamer. Pro 3D rekonstrukci se využívají dva odlišné přístupy - triangulace významných bodů a vytváření disparitní mapy. Tyto přístupy využívají různé metody, které se liší klíčovými vlastnostmi, kterými jsou především spolehlivost, přesnost a výpočetní čas. V současné době také dochází na tomto poli k velkému rozvoji metod založených na konvolučních neuronových sítích (CNN), které překonávají stávající metody v některých vlastnostech.

Rychlý vývoj na poli počítačového vidění přináší stále nové metody a přístupy pro zpracování dat z pasivního stereopáru za účelem 3D rekonstrukce scény. Účelem práce je vytvoření aktuálního přehledu těchto metod a jejich srovnání. Na základě znalosti funkce jednotlivých metod a jejich srovnání lze určit optimální metodu pro konkrétní aplikaci. Práce také obsahuje jednoduchý praktický příklad demonstrující použití rozebíraných přístupů a metod.

2 ANALÝZA PROBLÉMU A CÍL PRÁCE

2.1 Analýza problému

V současné době dochází k rychlému rozvoji přístupů a nových metod pro rekonstrukci 3D scény pomocí pasivního stereopáru. Tyto přístupy a metody se liší z hlediska mnoha klíčových vlastností, které zásadně ovlivňují jejich použitelnost. Existují 2 hlavní přístupy: triangulace významných bodů a vytváření disparitní mapy. Způsob, jakým je dosaženo výsledku, je u obou přístupů rozdílný.

Při triangulaci významných bodů jsou hledány jednoznačně geometricky určitelné body, které jsou popsány. Korespondující si body z dvojice snímků, které odpovídají stejnému bodu v prostoru, si jsou na základě popisu vzájemně přiřazeny. Následně se znalostí parametrů obou kalibrovaných kamer lze vypočítat přesnou 3D polohu těchto bodů v prostoru. V rámci tohoto přístupu existuje mnoho rozdílných metod pro nalezení a popis těchto významných bodů. Algoritmy pro hledání významných bodů se liší typem hledaných geometrických primitiv a matematickým popisem těchto míst. Algoritmy pro popis využívají blízké okolí významného bodu a liší se způsobem, jakým popisují vybrané okolí. Pro zmíněné úkony jsou stále častěji využívány také metody založené na konvolučních neuronových sítích. Neuronové sítě jsou díky své struktuře velmi vhodné pro operace s obrazovými daty.

Disparitní mapa je určena posunutím jednotlivých pixelů mezi dvěma obrazy. Čím blíže je objekt ke kamerám, tím je větší disparita (posunutí) mezi obrazy. Existují různé metody pro tvorbu disparitní mapy, které se liší výpočtem pro určení výsledné disparity.

2.2 Cíl práce

Práce podává přehled o vlastnostech přístupů 3D rekonstrukce scény pomocí pasivního stereopáru kamer a jejich metodách. Funkce rekonstrukce je demonstrována na jednoduchém praktickém příkladu.

Hlavním cílem je vytvoření kritického přehledu současného stavu poznání v oblasti zpracování snímků z pasivního stereopáru kamer pro 3D rekonstrukci objektu.

Dílčí cíle bakalářské práce:

- zpracovat přehled metod rekonstrukce 3D scény z páru fotografií
- porovnat metody z hlediska jejich klíčových vlastností
- vytvořit jednoduchý praktický příklad využití některé z těchto metod

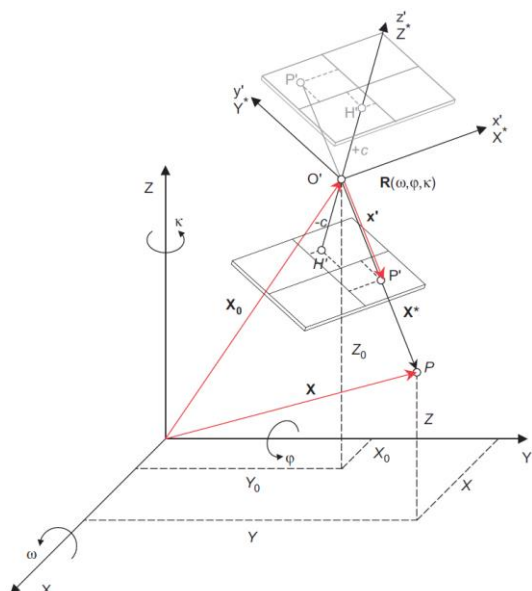
3 PŘEHLED SOUČASNÉHO STAVU POZNÁNÍ

Pasivní stereovize je postup získání prostorových dat na základě dvojice 2D snímků z kamer bez použití aktivního zdroje světla či jiného zařízení. Princip stereovize je podobný jako princip lidského prostorového vidění. Vzhledem k rozdílné poloze kamer jsou při snímání stejné scény rozdílné také vytvořené snímky. Scéna vhodná pro rekonstrukci pomocí stereopáru kamer musí být dostatečně strukturovaná.



Obr. 3-1 Pasivní stereopár kamer [1].

Předpokladem k 3D rekonstrukci scény z obrazu je kalibrace kamery. Kalibrace kamery je proces, během něhož se určují parametry transformace mezi souřadným systémem snímané scény a souřadným systémem čipu kamery (obrazu). Kalibrace probíhá tak, že je snímán etalon, kterým může být např. kalibrační šachovnice (Zhangova metoda), kalibrační tyče nebo kalibrační značky. Na základě znalosti polohy těchto bodů nebo vztahu mezi nimi lze vypočítat parametry kamery. Parametry kamery se dělí na interní a externí. Interní parametry určují vlastnosti optické soustavy kamery. Mezi interní parametry patří ohnisková vzdálenost, poloha hlavního snímkového bodu a koeficienty zkreslení kamery. Externí parametry vyjadřují polohu a natočení kamery v prostoru vzhledem k výchozímu bodu.



Obr. 3-2 Zobrazení parametrů kamery [2].

Na **Obr. 3-2** lze vidět schéma scény zachycené kamerou, kde dochází k zobrazení bodu P v prostoru na jeho obrazu P' . Zobrazení se řídí následujícími rovnicemi transformace mezi globálním souřadným systémem a souřadným systémem snímku (rovnice 1):

$$x' = x'_0 + z' \frac{r_{11}(X - X_0) + r_{21}(Y - Y_0) + r_{31}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)} + \Delta x'$$

$$y' = y'_0 + z' \frac{r_{11}(X - X_0) + r_{21}(Y - Y_0) + r_{31}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)} + \Delta y'$$

Kde X, Y, Z jsou souřadnice zobrazovaného bodu v globálním souřadném systému, x', y', z' jsou souřadnice souřadného systému kamery a ω, ϕ, κ jsou rotace kolem příslušných os globálního souřadného systému do souřadného systému kamery.

Při snímání obrazu dvěma kamerami dochází k tomu, že jednotlivé obrazy z obou kamer jsou vzájemně posunuté. Existují 2 různé přístupy pro rekonstrukci 3D scény z obrazů pořízených stereopárem.

Prvním přístupem je **triangulace významných bodů**. Na základě posunutí jednotlivých bodů v obraze lze pomocí triangulace určit polohu těchto bodů v prostoru. Tento přístup 3D rekonstrukce scény může sloužit k vytváření modelů nebo měření rozměrů objektů. Jeho výpočet probíhá v několika fázích:

- Detekce významných bodů - nalezení a určení polohy významných bodů
- Přiřazení odpovídajících si párů významných bodů - vytvoření popisu a korespondence
- Triangulace odpovídajících si bodů - zjištění souřadnic

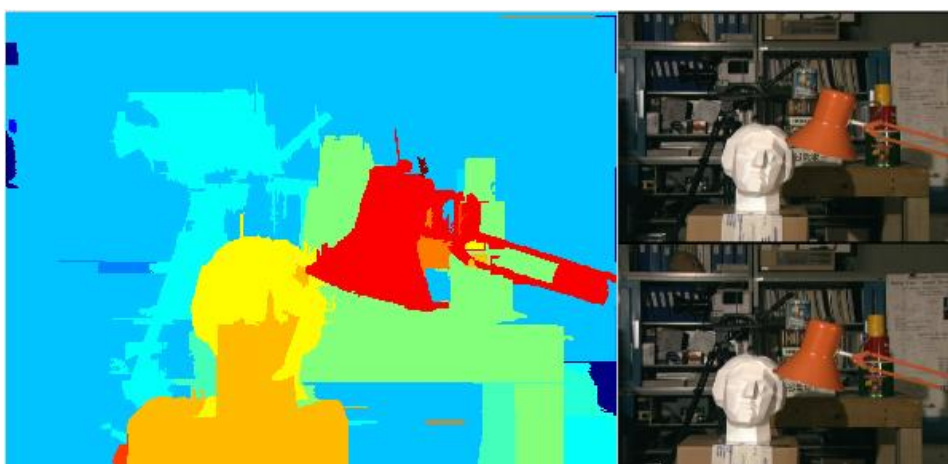


Obr. 3-3 Významné body (vlevo), korespondující si body ze dvou snímků (vpravo) [3].

Druhým přístupem je výpočet **disparitní mapy**. Disparitní mapa vyjadřuje rozdíl pozic stejného bodu na snímcích scény pořízených stereopárem. Čím blíže od kamer se objekt nachází, tím větší je posunutí mezi snímky. Výpočet disparitní mapy probíhá v těchto fázích:

- Rektifikace - transformace souřadnic jednoho snímku do souřadnic druhého snímku
- Výpočet disparity - výpočet podobností obrazu pro jednotlivé pixely
- Finální úprava - výběr nejpodobnějších pixelů a korekce výsledků

Výsledek může být znázorněn například barevnou stupnicí reprezentující hloubku scény. Jelikož porovnávání jednotlivých pixelů nevede k uspokojivým výsledkům, algoritmy většinou porovnávají celé skupiny pixelů s předpokladem, že se na snímku nacházejí větší plochy s podobnou hloubkou. Problém této metody mimo jiné představují oblasti, které jsou na jednom obrázku zakryté za hranou objektu. V těchto oblastech vznikají nespojitosti disparitní mapy. Rychlý odhad hloubky scény může sloužit k navigaci robotů autonomních automobilů nebo pro aplikace jako je virtuální nebo rozšířená realita (VR/AR).



Obr. 3-4 Disparitní mapa a originální obrazy scény [4].

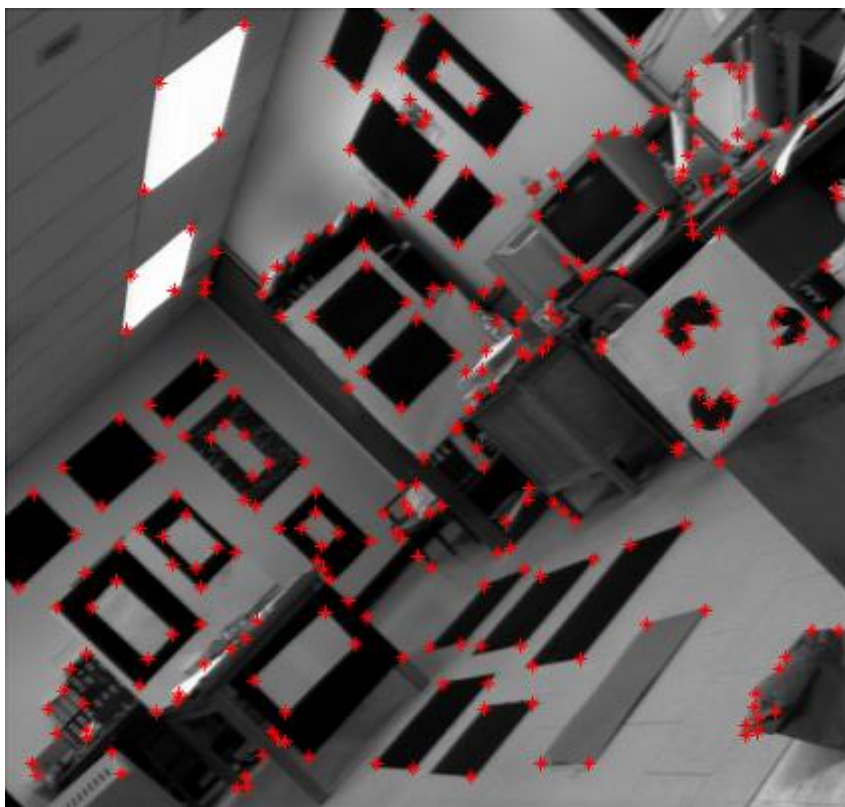
3.1 3D rekonstrukce scény pomocí triangulace významných bodů

3.1.1 Detektory významných bodů

Významný bod (klíčový bod, bod zájmu, „Feature“) je „zajímavé“ místo v obraze, nejčastěji osamocené body, hrany, rohy, významné oblasti. Tato místa jsou jasné matematicky a geometricky definovatelná. Významné body jsou důležité při hledání korespondencí mezi více obrazy, což je nutnou součástí 3D rekonstrukce scény. Významný bod má jasné daný matematický popis, aby byly možné opakované detekce.

Existuje velké množství detektorů, které slouží k hledání významných bodů s různými vlastnostmi. Vhodnost použití konkrétního detektoru závisí na požadovaných vlastnostech. Invariance detektoru na obrazové změny zaručují stabilní významné body (invariance detektoru vůči natočení, zkosení, změně měřítka a dalším transformacím obrazu nebo invariance na šum, rozmazání či změnu osvětlení apod.).

Další důležitou vlastností detektoru je rychlost detekce. Vysoká rychlost detekce je požadována především u aplikací fungujících v reálném čase, jako je například rozšířená realita nebo navigace autonomních automobilů či robotů.



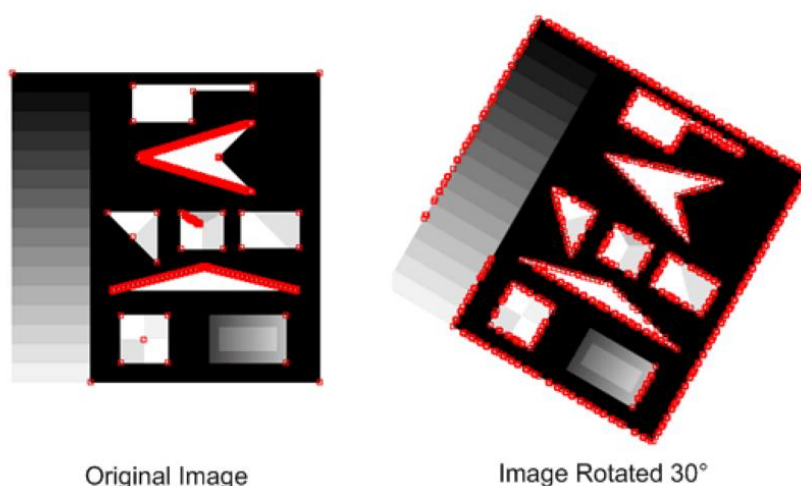
Obr. 3-5 Významné body nalezené detektorem rohů [5].

Moravcův detektor

Moravcův detektor měří rozdíly intenzity jasu v malé čtvercové oblasti, která je posouvána v několika směrech [6]. Mohou nastat tyto tři možnosti:

- Když je oblast rovná, posuvy ve všech směrech mají malou změnu intenzity.
- Když oblast narazí na hranu, posun rovnoběžný s hranou bude mít malou změnu intenzity, ale posuv kolmý k hraně velkou změnu intenzity.
- Když oblast narazí na roh nebo osamocený bod, nastane velká změna intenzity jasu ve všech směrech posuvu.

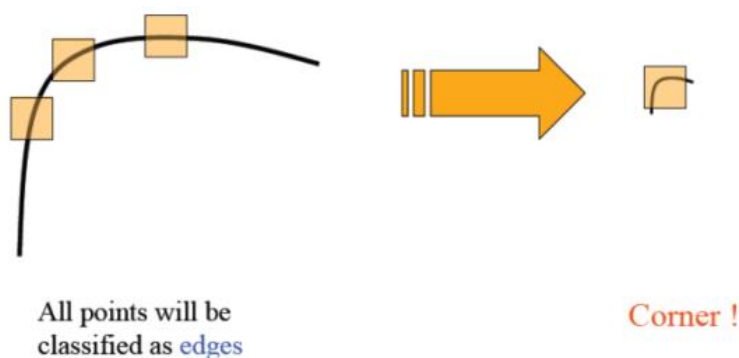
Tento detektor je jednoduchý na výpočet, ale má několik nedostatků. Kvůli tomu, že jednotlivé posuvy svírají úhel 45° , při otočení obrazu jsou detekovány různé významné body (viz **Obr. 3-6**) a často vyhodnotí hranu jako významný bod. Odezva detektoru je náchylná na šum kvůli tomu, že sledovaná oblast je čtvercová. Dále detektor není odolný na změnu měřítka.



Obr. 3-6 Zobrazení náchylnosti Moravcova detektoru na rotaci [7].

Harris-Stephensův detektor

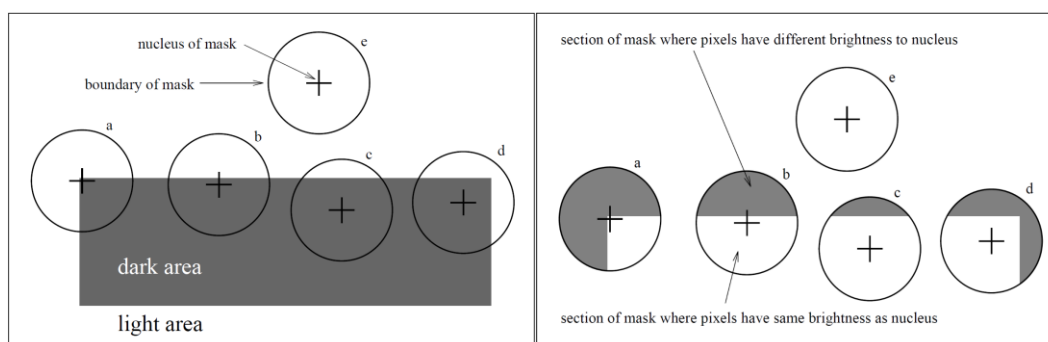
Jedná se o vylepšený Moravcův detektor [8]. Byl vyřešen problém diskrétních posuvů pomocí autokorelační funkce, která může měřit změnu intenzity v libovolných směrech. Díky tomu se zlepšila přesnost lokalizace významných bodů a přesnost při rotaci obrazu. Sledovaná oblast je tzv. Gaussova funkce, která udává váhu jednotlivých pixelů v závislosti na vzdálenosti od vyšetřovaného bodu. Toto opatření zvýšilo odolnost detektoru vůči šumu. Nevýhodou je však delší čas výpočtu ve srovnání s Moravcovým detektorem a hlavně fakt, že je závislý na změně měřítka, kdy při přiblížení obrazu nemusí být některé významné body detekovány (viz **Obr. 3-7**).



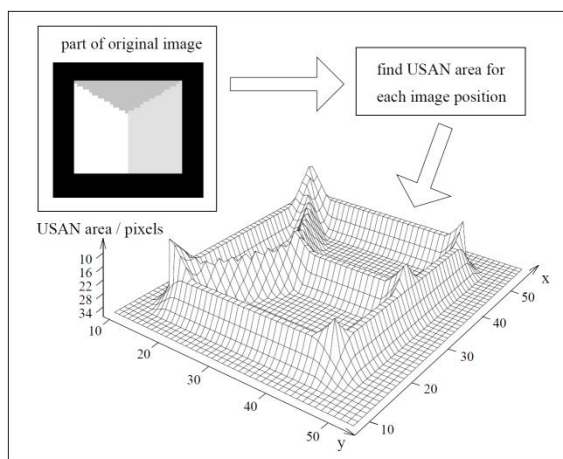
Obr. 3-7 Náchylnost Harrisova detektoru na změnu měřítka [7].

SUSAN

Metoda SUSAN (Smallest Univalued Segment Assimilating Nucleus) je prezentována v [9]. Tento detektor porovnává jas každého pixelu uvnitř kruhové oblasti s prostředním (vyšetřovaným) pixelem. Následně části kruhové oblasti, které mají stejný nebo podobný jas jako vyšetřovaný bod, jsou označeny jako USAN (Univalued Segment Assimilating Nucleus). Velikost USAN je největší, pokud sledovaná oblast má všude stejný jas jako vyšetřovaný bod, poloviční, pokud je jádro blízko hrany a minimální uvnitř rohu. SUSAN vyhodnotí body s lokálně minimální hodnotou USAN jako významné body. Změnou velikosti mezní hodnoty USAN lze ovlivnit, zda se bude jednat o detektor rohů nebo detektor hran. Výhodou tohoto detektoru je rychlost výpočtu, dobrá odolnost vůči rotaci a změně osvětlení. Nevýhodou je citlivost na šum.



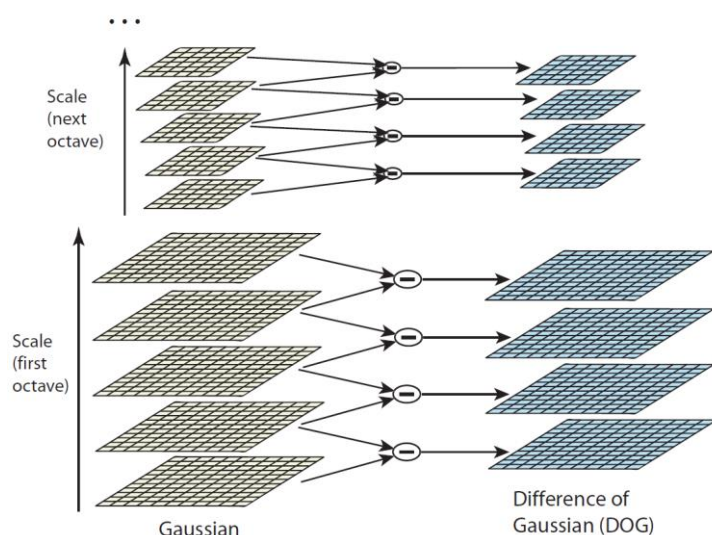
Obr. 3-8 Pět kruhových oblastí s vyznačeným středem různě umístěných na jednoduchém obraze (vlevo). Pět stejných kruhových oblastí. Znáznornění oblasti USAN bílou barvou (vpravo) [9].



Obr. 3-9 Trojrozměrný graf zobrazující velikosti oblasti USAN v závislosti na pozici vyšetřovaného bodu na obraze [9].

SIFT

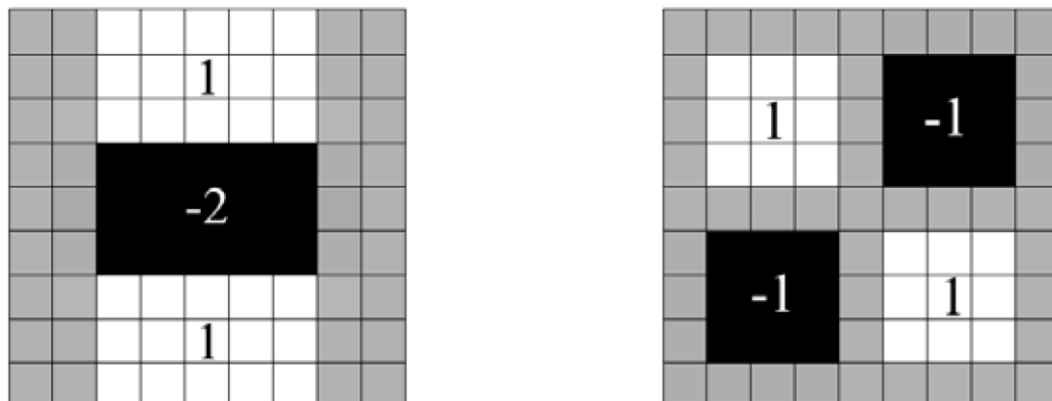
Metoda SIFT (Scale-Invariant Feature Transform) je prezentována v [10]. Jedná se o algoritmus, který spojuje funkci detektoru významných bodů a zároveň deskriptoru významných bodů (SIFT jako deskriptor významných bodů je uveden v kapitole 3.1.4). Detektor funguje na principu hledání lokálních extrémů funkce DoG (Difference of Gaussians). Jedná se o rozdíl mezi obrazy s různými hodnotami rozostření vytvořeného pomocí Gaussovy funkce. Dále je vytvořena tzv. Gaussovská pyramida, kdy je sestaveno několik měřítek obrazu. Proces hledání lokálních extrémů je zopakován pro všechny měřítka obrazu. Následuje post-processing, pomocí kterého jsou vyřazeny méně stabilní body, které mají nízký kontrast nebo leží na hranách. Body, které splnily všechny podmínky, jsou vybrány jako významné body. Hlavní výhodou tohoto detektoru je invariance vůči změně měřítka.



Obr. 3-10 Tvorba scale space a výpočet DoG [10].

SURF

Metoda SURF (Speeded-Up Robust Features) je prezentována v [11]. Jedná se o algoritmus, který je zároveň detektor významných bodů a deskriptor významných bodů (SURF jako deskriptor významných bodů v kapitole 3.1.4). Metoda SURF významné body detekuje pomocí determinantu Hessovy matice (hessián). Výpočet Hessovy matice je tedy konvoluce obrazu se čtvercovými filtry znázorněnými na **Obr. 3-11**. Aby nalezené významné body byly invariantní ke změně měřítka, je tento proces zopakován pro různá měřítka obrazu jako u metody SIFT. Pokud má daný bod ve svém okolí extrémní hodnotu, je považován za významný bod.



Obr. 3-11 Druhé derivace Gaussovy funkce podle y a podle xy s použitím čtvercových filtrů. (šedá pole mají hodnotu 0) [11].

KAZE

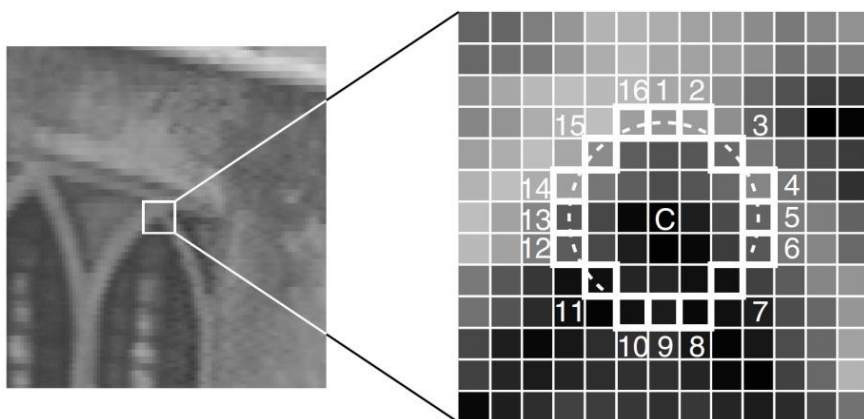
Opět se jedná o algoritmus [12], který v sobě spojuje jak detektor, tak deskriptor významných bodů (KAZE jako deskriptor významných bodů v kapitole 3.1.4). Tato metoda je podobná metodě SIFT a SURF. Na rozdíl od těchto metod využívá nelineárního difuzního filtru namísto Gaussovy funkce. Pro detekci významného bodu počítá odezvu na determinant Hessovy matice za různých měřítek. Body, které mají ve svém okolí extrémní hodnotu, jsou považovány za významné body.



Obr. 3-12 Porovnání Gaussova filtru (nahore) a nelineárního difuzního filtru (dole) pro různá zkraslení [12].

FAST

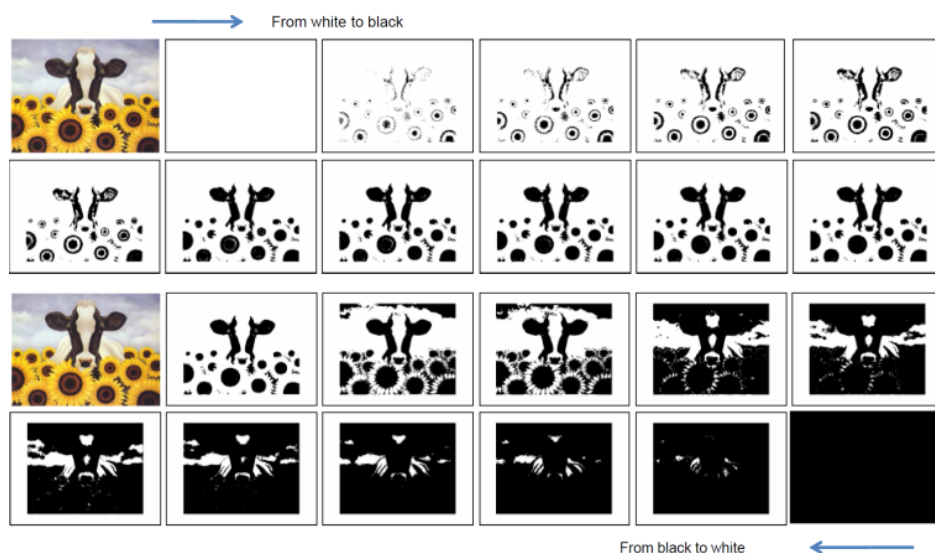
Metoda FAST (Features From Accelerated Segment Test) je prezentována v [13]. FAST je detektor významných bodů založený na porovnávání hodnot jasu vyšetřovaného pixelu oproti pixelům ležících na kružnici. Tato kružnice má poloměr 3 pixely, střed ve vyšetřovaném bodě a skládá se celkem z 16 pixelů zvolených pomocí Bresenhamova algoritmu. Pokud alespoň 12 vzájemně sousedících pixelů má o určitou prahovou hodnotu vyšší nebo nižší jas jak prostřední pixel, je středový pixel vyhodnocen jako významný bod. Rychlost algoritmu lze optimalizovat tím, že se netestují pixely v takovém pořadí, v jakém leží na kružnici, ale testují se diagonálně (nejprve body 1,9,5 a 13 viz **Obr. 3-13**). Takto lze zjistit, že se nejedná o významný bod mnohem rychleji, protože minimálně tři z těchto čtyř bodů musí splňovat podmínku (mít o určitou prahovou hodnotu vyšší nebo nižší hodnotu jasu než jas středového bodu), jinak se o významný bod nemůže jednat. Touto optimalizací se sníží počet testovaných pixelů z šestnácti na v průměru 3,8 pixelu. Výhodou tohoto detektoru je v první řadě rychlost, které dosahuje a velké množství nalezených bodů.



Obr. 3-13 Bresenhamova kružnice s 16 vyznačenými pixely a středem (c), body na pozicích 1,9,5 a 13 jsou vyšetřovány jako první [13].

MSER

Metoda (the Maximally Stable Extremal Regions) je prezentována v [14]. Jedná se o detektor založený na hledání stabilních oblastí. Detektor postupně provádí prahování obrazu pro všechny hodnoty intenzity jasu (mění prahovou hodnotu intenzity). S rostoucí prahovou hodnotou přibývají černé oblasti. Postupně dochází k propojování sousedních oblastí. Regiony neměnicí se přes větší rozsah prahových hodnot, jsou hledané maximálně stabilní regiony.



Obr. 3-14 Znáznornění prahování obrazu [15].

Detektor nejlépe funguje, pokud obraz obsahuje homogenní regiony, které mají jasné hranice. Na rozmazaných snímcích nedosahuje dobrých výsledků a je také náchylný na přirozenou změnu osvětlení, jako jsou například stíny.



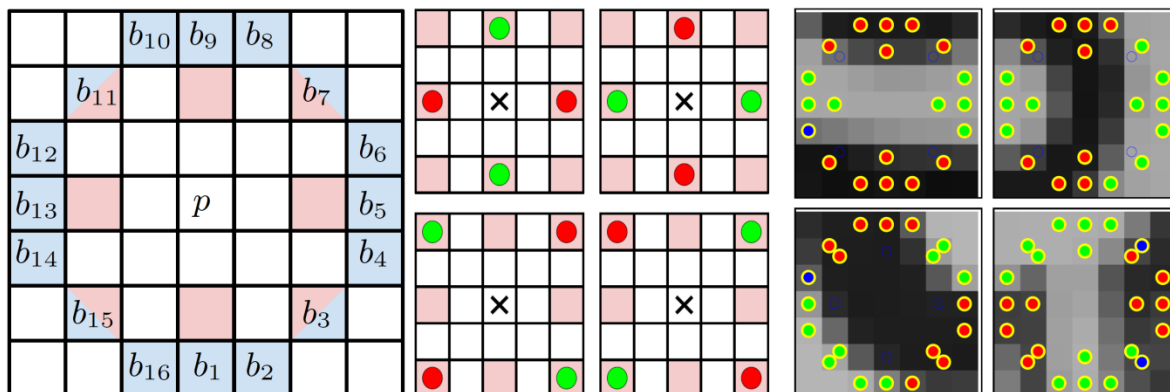
Obr. 3-15 Zobrazení oblastí nalezených pomocí MSER detektoru [14].

In the Saddle

Jedná se o detektor významných bodů, které se nacházejí v oblastech, jejichž poloha představuje styk dvou rohů nebo jakýkoliv jiný sedlo připomínající profil [16]. Detektor pracuje ve třech fázích. V první fázi detektor co nejrychleji eliminuje co nejvíce bodů tím, že testuje jas pixelů ležících na vnitřní kružnici okolo vyšetřovaného bodu, a to ve dvou párech na sebe kolmých směrů (x , y) viz **Obr. 3-16**. Test je úspěšný, pokud dva body jednoho směru jsou výrazně jasnější než body ve směru kolmém. Všechny čtyři možné případy jsou znázorněny na **Obr. 3-16**. Tímto se eliminuje okolo 80-85% všech bodů.

V druhé fázi detektor pracuje s 16 pixely Bresenhamovy kružnice okolo vyšetřovaného bodu na obrázku **Obr. 3-16** modré body. Jas každého pixelu ležícího na této kružnici je porovnán s jasnem prostředního pixelu a na základě toho jsou tyto pixely zařazeny do tří skupin. Světlejší než vyšetřovaný bod (l), podobné jako vyšetřovaný bod (s) a tmavší než vyšetřovaný bod (d). Test je úspěšný, pokud se na kružnici nachází přesně čtyři oblouky skupin (l) a (d) v alternujícím pořadí. Každý oblouk musí obsahovat 2-8 pixelů a mezi každým rozhraním dvou skupin (l) a (d) se mohou nacházet až dva pixely skupiny (s).

Poslední fází je finální úprava, kdy jsou potlačeny body v blízkém okolí nalezeného významného bodu tak, aby jeden významný bod nebyl vyhodnocen vícekrát.



Obr. 3-16 Znáznornění Bresenhamovy kružnice (modré pixely) a vnitřní kružnice (růžové body) (vlevo). Čtyři přípustné kombinace jasu pixelů z první fáze detekce (uprostřed). Příklady detektorem vyhodnocených významných bodů (vpravo). Barvy odpovídají skupinám: světlejší oblast (l) - zelená, tmavší oblast (d) - červená, oblast podobného jasu (s) - modrá [16].

3.1.2 CNN detektory významných bodů

Konvoluční neuronové sítě (Convolutional Neural Network, CNN) jsou speciálním druhem klasických neuronových sítí, které jsou díky své struktuře vhodné pro zpracování obrazových dat. Podstatou konvolučních neuronových sítí jsou konvoluční vrstvy. Jednotlivé vrstvy mají pevnou strukturu propojení. Většina výpočtů v rámci CNN je uskutečněna pomocí grafického procesoru (GPU), který je ve specifických výpočtech výkonnější, než hlavní procesor počítače (CPU) a neustále dochází k zvyšování jejich výkonu a rychlosti. Jedna z hlavních nevýhod CNN je trénování, ke kterému je potřeba velké množství dat, a je tak výpočetně i časově náročné. Komplikované sítě se na nejnovějších GPU trénují několik dnů až týdnů. Trénink CNN lze rozdělit na dva typy:

- učení s učitelem (supervised learning), kde pro vstupní data je určen správný výstup
- učení bez učitele (unsupervised learning), kde ke vstupním datům není známý výstup

Metody založené na CNN se během několika let dokázaly vyrovnat konvenčním metodám detekce významných bodů a lze předpokládat, že dojde k zlepšení jak rychlosti výpočtů, tak spolehlivosti těchto metod. Kvůli zmíněnému trénování je aplikace CNN náročnější než aplikace konvenčních metod detekce. Jednotlivé CNN bývají trénované na různých datasetech. Trénování probíhá tak, že CNN upravuje svoje parametry, aby napodobovala konvenční metody detekce. Obecně na netrénovaných obrazech CNN nedosahují tak dobrých výsledků. Kvůli těmto nedostatkům nejsou tolik používané jako konvenční metody detekce významných bodů.

TILDE

Metoda TILDE (A Temporally Invariant Learned DEtector) je prezentována v [17]. Jedná se o detektor, který detekuje významné body i za velmi zhoršených podmínek, jako je např. změna osvětlení nebo počasí scény. CNN je trénována na snímcích pořízených ze stejného místa s předem určenými významnými body. Na trénovaných snímcích dosahuje mnohem lepších výsledků ve srovnání s konvenčními metodami. Na snímcích úplně odlišných od trénované sady nedosahuje takového výkonu, ale i přesto je s konvenčními metodami srovnatelný.

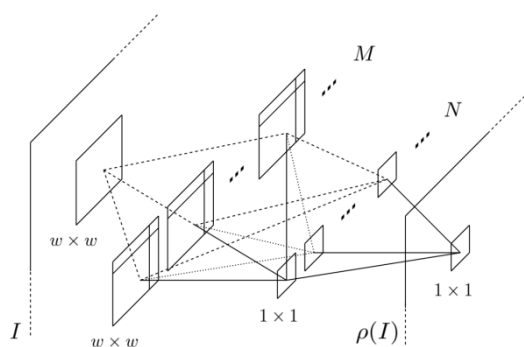
LIFT

Metoda LIFT (Learned Invariant Feature Transform) je prezentována v [18]. U této metody je detekce provedena pomocí po částech lineárních konvolučních filtrů. Získané body jsou robustní a odolné vůči změně jasu. Dobrých výsledků je u této metody dosaženo pouze na trénovaných obrazech a nalezené body nejsou invariantní na změnu měřítka a úhlu pohledu.

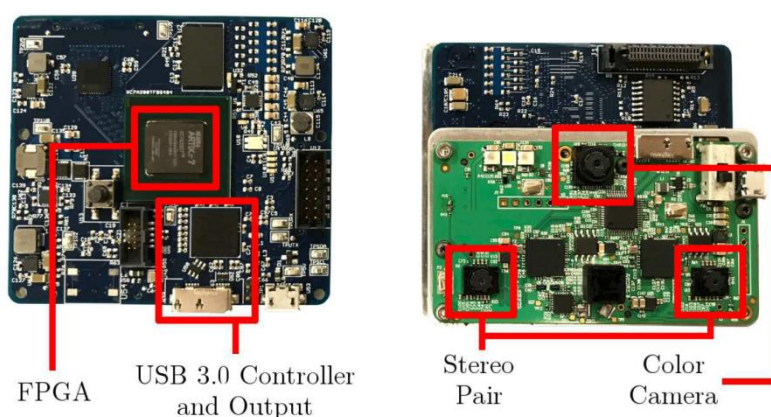
KCNN

Metoda KCNN (Extremely-Efficient Hardware Keypoint Detection with a Compact Convolutional Neural Network) je prezentována v [19]. Jedná se o velmi efektivní metodu detekce významných bodů pomocí CNN. Neuronová síť je složena ze tří kompaktních vrstev. Při trénování neuronová síť upravuje svoje parametry tak, aby napodobovaly konvenční metody detekce významných bodů a zjistilo se, že navržená architektura je schopna dosáhnout srovnatelných výsledků. Struktura CNN umožňuje, že může být jednoduše přetrénována k detekování bodů námi potřebných vlastností. Největší výhodou této metody je rychlost. Nejlepších rychlostí metoda dosahuje při realizaci na FPGA čipu (Field Programmable Gate Array - programovatelné hradlové pole). Rychlost detekce se v takovém případě pohybuje okolo časů menších než jedna milisekunda na obraze o rozlišení 1280×800 , oproti běžným časům, které se pohybují v desítkách až stovkách milisekund, při zachování ostatních vlastností srovnatelných s konvenčními metodami. Tato metoda díky své extrémně nízké latenci může být použita u aplikací probíhajících v

reálném čase jako je například AR / VR, autonomní řízení a díky malým rozměrům a nízké spotřebě energie FPGA čipů, také v integrovaných systémech.



Obr. 3-17 Architektura sítě KCNN. I je vstupní obrázek a $\rho(I)$ je výsledná odezva významných bodů. Každé okno znázorňuje konvoluční operaci. Filtry v první konvoluční vrstvě jsou oddělitelné [19].

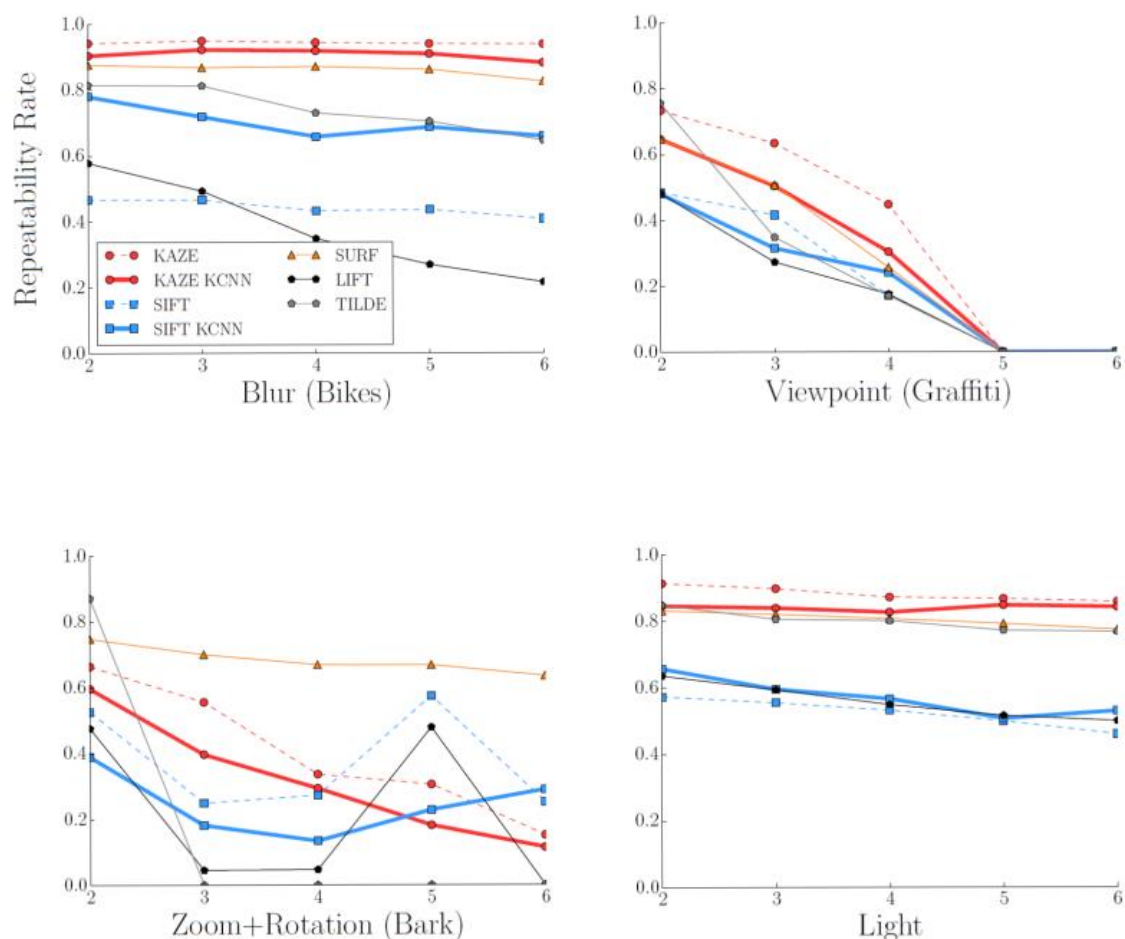


Obr. 3-18 Znáznornění kompaktního hardware systému, na kterém je implementován KCNN [19].

3.1.3 Srovnání detektorů

U detektorů je nejdůležitější vlastností stabilita nalezených významných bodů, tedy schopnost detektoru detekovat stejný významný bod za různých podmínek. Nejčastěji se detektory porovnávají na invarianci vůči rotaci, změně měřítka, osvětlení nebo úhlu pohledu. Další důležitou vlastností je počet nalezených významných bodů, kdy obecně platí čím více nalezených významných bodů, tím lépe. Především pro real-time aplikace je také velmi důležitá rychlost výpočtu, kterou ovlivňuje složitost a optimalizace výpočtu. Některé detektory jako jsou Moravcův, Harris-Stephensův a SUSAN se v dnešní době téměř nepoužívají. Pro svoji jednoduchost a názornost mohou sloužit jako ilustrace principu detekce významných bodů.

V článku [19] je proveden test srovnávající vybrané detektory. Pro testování byly použity snímky z Oxfordského datasetu [6]. Detektory byly testovány z hlediska stability nalezených významných bodů. Dále byly detektory upraveny tak, aby počet nalezených významných bodů byl u všech metod srovnatelný.



Obr. 3-19 Zobrazení invariance jednotlivých metod na rozmazání, změnu úhlu pohledu, měřítka s rotací a osvětlení. Vertikální osa znázorňuje míru opětovného nalezení bodu. Horizontální osa znázorňuje míru zkreslení [19].

Z hlediska stability nalezených významných bodů vychází nejlépe metoda KAZE, která má nejlepší invarianci vůči rozmazání, změně úhlu pohledu a změně osvětlení. Následuje konvoluční neuronová síť KCNN, která je trénovaná právě podle metody KAZE. Dále je vidět, že při současné rotaci a změně měřítka obrazu vychází nejlépe metoda SURF. Při zachování stejných podmínek, jako byly u testu stability, je porovnána metoda KAZE s metodou KCNN z hlediska doby potřebné k detekci významných bodů na obraze o rozlišení 1280×800 .

tab. 1 Zobrazení doby detekce významných bodů pro metody KAZE a KCNN na různých procesorech [19].

Metoda	Procesor	Latence / ms
KAZE	Intel I7-5930K (CPU)	58
KCNN	Intel I7-5930K (CPU)	289
KCNN	NVIDIA GTX 1080 (GPU)	12
KCNN	Xilinx Zynq 7020 (FPGA)	1

Mnohonásobně nižších výpočetních časů dosáhla metoda KCNN při realizaci na FPGA čipu. Právě při správné optimalizaci lze pomocí konvolučních neuronových sítí zkrátit výpočetní čas potřebný pro nalezení významných bodů.

Další test z článku [20] srovnává vybrané detektory významných bodů z hlediska počtu nalezených bodů a výpočetního času. Pro testování bylo využito dvou datasetů. Jeden dataset byl vytvořen z několika snímků popsaných v článku [20], druhý pochází ze snímků Oxfordské univerzity [6].

tab. 2 Srovnání počtu významných bodů nalezených vybranými detektory (průměr ze dvou datasetů) [20].

Metoda	Počet nalezených významných bodů
SIFT	3126
SURF	4318
KAZE	1517
FAST	9782

Největší počet významných bodů byl nalezen pomocí detektoru FAST. Naopak nejméně významných bodů bylo nalezeno pomocí detektoru KAZE. Mezi jednotlivými metodami jsou násobné rozdíly.

3.1.4 Deskriptory významných bodů

Deskriptor popisuje okolí konkrétního významného bodu. V případě rekonstrukce scény pomocí pasivního stereopáru se deskriptory používají k určování odpovídajících si párů významných bodů. Existuje velké množství deskriptorů s různými vlastnostmi a vhodnost použití konkrétního detektoru závisí na požadovaných vlastnostech pro danou aplikaci.

Základní vlastností deskriptoru je schopnost dostatečně popsat okolí významného bodu tak, aby byla možná korespondence bodů z různých obrázků. Zároveň deskriptor musí být dostatečně kompaktní z důvodu menších nároků na paměť a výpočet. Dále by měl být odolný vůči změně geometrie (natočení, zkosení, změna měřítka nebo jiné transformace obrazu) a osvětlení. Čím lepší jsou tyto vlastnosti, tím spolehlivější bude nalezení shod

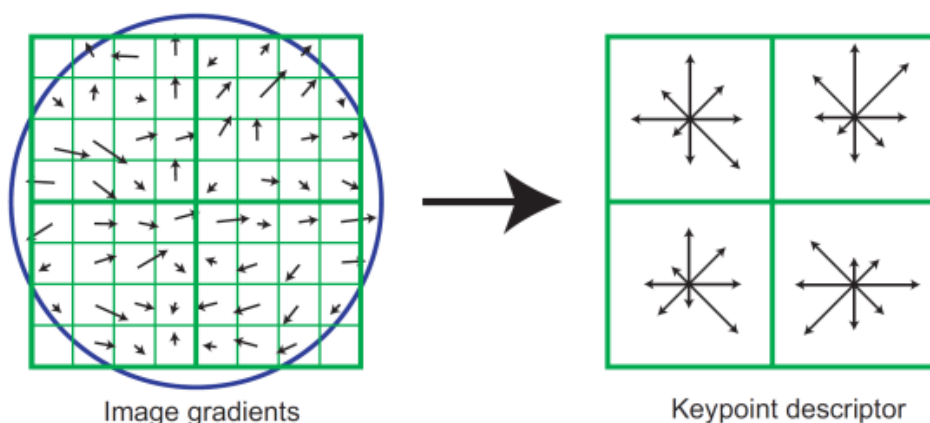
mezi jednotlivými obrazy. Další důležitou vlastností deskriptoru je rychlost, která limituje některé detektory v rozsahu použití.

SIFT

Metoda SIFT (Scale Invariant Feature Transform - Transformace významných bodů nezávislých vůči měřítku) je prezentována v [10]. Jedná se o algoritmus, který je zároveň detektor významných bodů a deskriptor významných bodů (SIFT jako detektor významných bodů v kapitole 3.1.1).

Ke každému významnému bodu je přiřazena orientace, aby bylo dosaženo invariance na rotaci. V okolí významného bodu jsou vypočítány velikosti gradientu jasu a celá oblast je rozdělena na 36 podoblastí ve tvaru kruhových výsečí po deseti stupních. V rámci každé podoblasti jsou gradienty sečteny a je vytvořen orientační histogram. Nejvyšší hodnota v histogramu představuje výslednou orientaci. Pokud některá hodnota v histogramu dosahuje 80% nejvyšší hodnoty, je také použita k určení orientace a vytvoří se významný bod se stejným umístěním a měřítkem, ale různým vektorem orientace. To přispívá k větší stabilitě významného bodu.

Deskriptor vhodně popisuje okolí významného bodu o určité velikosti. Okolí zkoumaného bodu je rozděleno na 4×4 stejně velkých čtvercových oblastí a každá tato oblast je dále rozdělena na 4×4 podoblastí. Rozčlenění okolí do několika podoblastí má výhodu v tom, že výsledný deskriptor je odolný vůči malým posuvům obrazu. V každé podoblasti je spočítán gradient jasu a pro každou oblast je z těchto gradientů sestaven histogram orientací. Výsledný deskriptor se skládá z šestnácti histogramů nacházejících se v okolí významného bodu a je označován jako 128-binový deskriptor (okolí obsahuje 4×4 oblastí a každá obsahuje 8-binový histogram). Tento deskriptor je odolný vůči změně měřítka a jasu.



Obr. 3-20 Zjednodušená ukázka deskriptoru SIFT [10].

SURF

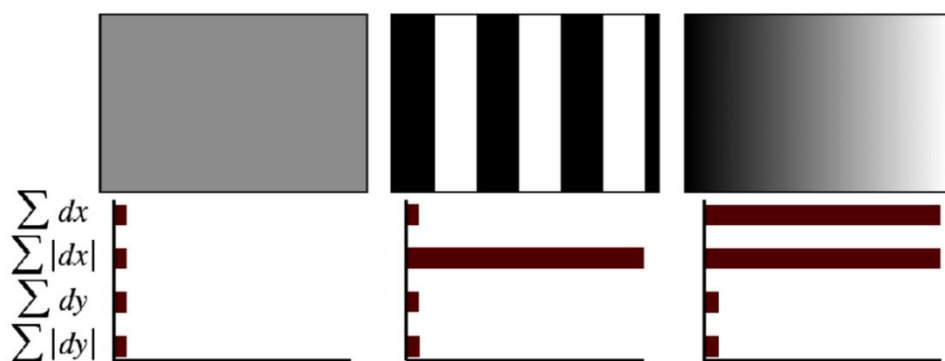
Metoda SURF (Speeded-Up Robust Features) je prezentována v [11]. Jedná se o deskriptor i detektor významných bodů (SURF jako detektor významných bodů v kapitole 3.1.1), jehož hlavním cílem byla konstrukce výpočetně rychlé a stabilní metody na deskripci významných bodů, která by byla schopna pracovat v reálném čase. Metoda využívá podobných principů jako SIFT. Zrychlení oproti SIFT je dosaženo v sestavení měřítkově nezávislé reprezentace obrazu a v redukci velikosti výsledných deskriptorů.

Podobně jako u SIFT je přiřazena každému významnému bodu orientace, aby výsledný deskriptor byl invariantní na rotaci. K tomu se využívá odezva obrazu na Haarovu vlnku v kruhovém okolí významného bodu. Toto kruhové okolí má střed ve významném bodě a poloměr závisí na měřítku významného bodu. Velikost Haarovy vlnky je také závislá na měřítku. Výsledky konvoluce obrazu s Haarovou vlnkou jsou upraveny pomocí Gaussovy funkce se středem ve významném bodě. Jsou zakresleny jako vektory, které jsou sečteny v rámci šesti kruhových výsečí okolo významného bodu. Na základě součtu s největší hodnotou je určen dominantní směr.



Obr. 3-21 Haarova vlnka (vlevo) zobrazení dominantní orientace (vpravo) [11].

Dalším krokem je určení popisu. V okolí významného bodu je vytvořena čtvercová oblast. Tato oblast má střed ve významném bodě a je natočena ve směru dominantní orientace. Dále je oblast rozdělena na 4×4 stejně velké podoblasti. V každé podoblasti je určeno pět pravidelně rozmístěných bodů, pro které je vypočtena jejich odezva na Haarovou vlnku, podobně jako při určování orientace. Výsledky konvoluce jsou opět upraveny pomocí Gaussovy funkce se středem ve významném bodu a zakresleny jako vektory. V rámci každé podoblasti jsou vektory sečteny a z důvodu větší odolnosti deskriptoru na změnu osvětlení jsou také spočítány součty těchto vektorů v absolutní hodnotě. Tím vznikne čtyřrozměrný vektor. Výsledný detektor se skládá z takto získaných vektorů pro všech 16 podoblastí a okolí významného bodu tak popisuje pomocí 64 binů.



Obr. 3-22 Odezva Haarovy vlny na různé vzory [11]

KAZE

Jedná se o deskriptor i detektor významných bodů [12] velmi podobný metodě SURF. Aby deskriptor byl invariantní vůči rotaci, je určena nejprve orientace deskriptoru. V první fázi se vytvoří kruhová oblast okolo významného bodu, která je rozdělena na 6 stejných podoblastí. Následně se spočítají derivace intenzity jasu všech bodů uvnitř kruhové oblasti ve směru osy X a Y . Derivace jsou upraveny pomocí Gaussovy funkce se středem ve významném bodu a jsou znázorněny jako vektory. Pro každou ze šesti podoblastí jsou sečteny tyto vektory a nejdelší z nich představuje výslednou orientaci.

V druhé fázi deskripce je vytvořena čtvercová oblast. Tato oblast je rozdělena na 4×4 podoblastí tak, že každá podoblast částečně přesahuje do sousední podoblasti. Poté jsou spočteny derivace intenzity jasu ve směru osy X a Y , které jsou upraveny pomocí Gaussovy funkce se středem ve významném bodu a rozptylem $1,5\sigma$. Takto upravené hodnoty derivací jsou sečteny v rámci jedné podoblasti a pro větší odolnost vůči změně osvětlení je vypočítán součet absolutních hodnot. Tak je vytvořen deskriptor popisující okolí významného bodu pomocí čtyřrozměrného vektoru pro 16 podoblastí o celkové velikosti 64 binů.

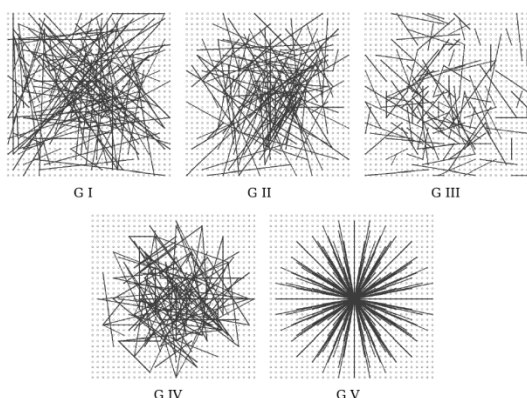
Metoda AKAZE (Accelerated-KAZE) je vylepšenou verzí detektoru KAZE [21]. Využívá numerické metody FED – Fast Explicit Diffusion, která zrychluje proces výpočtu polohy a popisu významného bodu. Zároveň je zachována přesnost a stabilita významných bodů.

BRIEF

Metoda BRIEF (Binary Robust Independent Elementary Features) je prezentována v [22]. Jedná se o deskriptor, který se snaží minimalizovat velikost paměti potřebné k popisu bodu tím, že nepoužívá čísla s plovoucí desetinnou čárkou. To sníží čas potřebný k vytvoření korespondencí bodů. Nevýhodou je fakt, že BRIEF není schopen vypočítat orientaci a při natočení obrazu nedosahuje dobrých výsledků.

Nejprve je vstupní obraz vyhlazen pomocí Gaussova filtru. Následuje test, který porovnává intenzity jasu dvou náhodných bodů. Výsledek může mít hodnotu 1, pokud první bod má vyšší intenzitu jasu než druhý, v opačném případě má výsledek hodnotu 0. Toto porovnání

je provedeno pomocí konkrétního počtu párů (128, 256, 512). Počet testovaných párů ovlivňuje rychlost výpočtu, množství uchovávaných dat a přesnost.



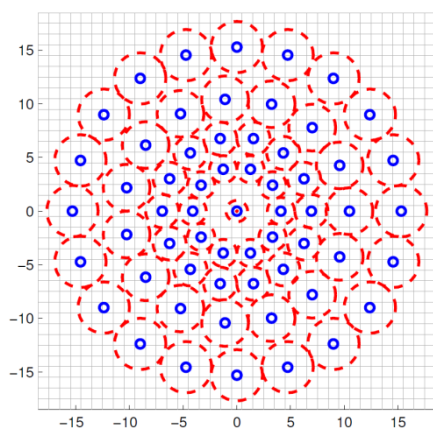
Obr. 3-23 Náhodně propojené dvojice bodů, symetricky propojené body (vpravo dole) [22].

BRISK

Metoda BRISK (Binary Robust Invariant Scalable Keypoints) je prezentována v [23]. Jedná se o deskriptor podobný metodě BRIEF, při které jsou porovnávány intenzity jasu náhodných párů bodů. U metody BRISK porovnávané body leží na soustředných kružnicích se středem ve významném bodu. Obrázek je v první fázi vyhlazen pomocí Gaussovy funkce. Podle vzdálenosti jsou páry rozděleny na vzdálené a blízké. Vzdálené body jsou použity pro odhad orientace a blízké body pro určení deskriptoru. Odhad orientace je proveden tak, že je určen gradient intenzity jasu mezi vzdálenými páry. Všechny tyto gradienty jsou sečteny a tvoří výslednou orientaci.

Aby byl deskriptor invariantní vůči rotaci, jsou blízké páry pootočený ve směru orientace. Následuje test, který porovná intenzitu jasu blízkých párů. Výsledek testu je 1, pokud první bod má vyšší intenzitu jasu než druhý, v opačném případě má výsledek hodnotu 0. Deskriptor má délku 512 binů.

Největší výhodou tohoto deskriptoru je rychlost. Výpočetní časy metody BRISK jsou řádově kratší než u metod SIFT nebo SURF.



Obr. 3-24 Znáznornění soustředných kružnic ležících okolo významného bodu [23].

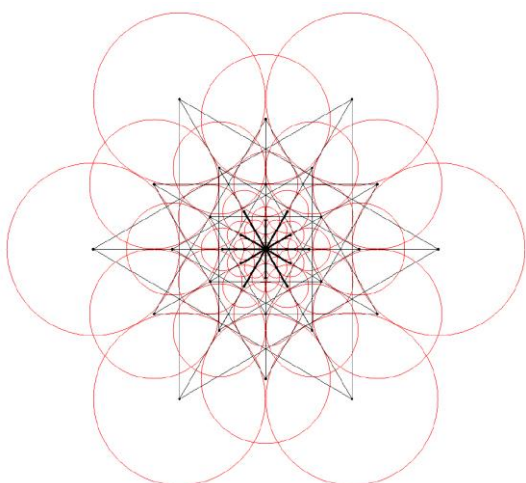
FREAK

Metoda FREAK (Fast Retina Keypoint) je prezentována v [24]. Jedná se o deskriptor podobný metodám BRIEF a BRISK, které porovnávají intenzity jasu párů bodů. Metoda FREAK je specifická v rozmístění porovnávaných bodů okolo významného bodu. Snahou tohoto deskriptoru je napodobit lidskou sítnici tím, že hustota porovnávaných bodů směrem od významného bodu klesá exponenciálně.

Pro určení orientace je použito 45 párů symetricky rozmístěných bodů viz **Obr. 3-25**. Mezi těmito páry jsou vypočítány gradienty intenzity jasu a součet všech gradientů určí výslednou dominantní orientaci.

Deskriptor je vytvořen pomocí 512 předem určených párů, následně proběhne test stejný jako u metody BRISK. Test má výsledek 1, pokud první bod má vyšší intenzitu jasu než druhý, v opačném případě 0.

Tato metoda je velmi rychlá a s nízkými nároky na paměť. Je tedy vhodná pro aplikace v reálném čase.



Obr. 3-25 Rozložení porovnávaných bodů použitých k určení orientace deskriptoru [24].

3.1.5 CNN deskriptory

CNN deskriptory bývají mnohem úspěšnější než konvenční, a to díky možným optimalizacím a změnám algoritmu. Pomocí velkého množství vstupních dat se může síť dokonce naučit rozpoznávat jednotlivé objekty. V posledních několika letech vznikla velká řada CNN deskriptorů. Jednotlivé metody se liší ve způsobu trénování, podle druhu a počtu použitých datasetů, které byly pro trénování použity. Jednotlivé metody založené na CNN jsou převážně trénovány pomocí konvenčních metod pro popis významných bodů.

LIFT

Metoda LIFT (Learned Invariant Feature Transform) je prezentována v [18]. Je metoda detekce a popisu významných bodů pomocí CNN trénovaných s učitelem (supervised CNN). Metodu lze rozdělit do tří vzájemně navazujících procesů: detektor významných bodů, odhad orientace a deskriptor významných bodů. Každý proces je založen na neuronových sítích (LIFT jako detektor v kapitole 3.1.2).

Deskriptor metody LIFT se trénuje pomocí obrazů s předem určenými korespondujícími si významnými body, které jsou otočeny ve směru dominantní orientace. CNN upravuje svoje parametry tak, aby minimalizovala rozdíly mezi popisy odpovídajících si bodů, a zároveň zvětšuje rozdíly mezi popisy neodpovídajících si bodů. Určení korespondujících si bodů a orientace je provedeno pomocí konvenčních metod.

Odhad orientace metody LIFT se trénuje pomocí obrazů s určenými korespondujícími si významnými body, které jsou nalezeny pomocí konvenčních metod detekce a popsány pomocí deskriptoru metody LIFT. CNN upravuje svoje parametry tak, aby minimalizovala rozdíl mezi popisem odpovídajících si natočených bodů na dvou snímcích stejné scény.

3.1.6 Srovnání deskriptorů

Nejdůležitější vlastností deskriptorů je schopnost popsat okolí významného bodu tak, aby byla možná následná korespondence bodů z různých obrázků. Tato vlastnost se nejčastěji hodnotí na základě poměru počtu správných korespondencí vůči počtu nalezených významných bodů (Matching score, MS) nebo počtu správných korespondencí vůči celkovému počtu korespondencí (Recall, RC). Dále je důležitá stabilita popisu, kdy výsledný popis musí být invariantní při různých obrazových změnách, jako je rotace, změna měřítka, osvětlení a úhlu pohledu. Pro aplikace v reálném čase je velmi důležitá rychlost, kterou ovlivňuje velikost popisu, kdy čím je popis kompaktnější, tím zpravidla bývá rychlejší výpočet popisu a doba potřebná ke korespondenci významných bodů. Testování probíhá na datasetech, které se liší v počtu snímků a jednotlivé snímky mají různé deformace obrazu, vůči kterým mají být deskriptory v praxi odolné (test na datasetu již zahrnuje, jak jsou algoritmy odolné vůči výše zmíněným deformacím).

V článku [21] je proveden test srovnávající vybrané deskriptory. Pro testování byly použity sady snímků z Oxfordského datasetu [6]. Deskriptory jsou srovnány pomocí matching score a recall. Jednotlivé deskriptory pracují s významnými body nalezenými pomocí vlastních detektorů (BRIEF a BRISK pomocí detektoru FAST).

tab. 3 Srovnání deskriptorů podle MS a RC [21].

Deskriptor	Dataset1		Dataset2	
	MS / %	RC / %	MS / %	RC / %
SIFT	8	69	15	62
SURF	27	68	11	52
BRIEF	25	67	7	17
BRISK	3	55	3	37
KAZE	37	81	16	54
AKAZE	42	76	11	33

Z prvního srovnání vychází nejlépe metoda KAZE a SIFT. Výsledky jednotlivých metod jsou na rozdílných datasetech úplně odlišné. Například metody AKAZE a BRIEF dosáhly na druhém datasetu polovičního výkonu jak na prvním datasetu. Mezi různými metodami dochází k násobným rozdílům ve stabilitě popisu.

Druhý test z článku [18] srovnává jednotlivé metody pomocí matching score. Celkově na třech různých datasetech ze článků [25], [26] a [17]. Deskriptory byly testovány na významných bodech nalezených podle následujících detektorů: SIFT, SURF, KAZE, BRIEF pomocí detektoru SIFT; následně BRISK a FREAK pomocí detektoru FAST; a pro deskriptor LIFT byl použit vlastní detektor.

tab. 4 Průměrné Matching score pro vybrané deskriptory [18].

	SIFT	SURF	BRIEF	BRISK	FREAK	KAZE	LIFT
Dataset1	28	20	16	20	18	25	37
Dataset2	27	24	13	19	19	21	32
Dataset3	13	18	12	12	12	20	19

Z druhého srovnání vychází nejlépe metoda LIFT, kdy na prvních dvou datasetech CNN překonává konvenční metody deskripce významného bodu. Důležité je zmínit, že první dva datasety jsou podobné těm, na kterých byla neuronová síť trénovaná. Také v článku neuvádí hodnotu recall, takže nelze zjistit počet špatně nalezených korespondencí. Na třetím datasetu je lift srovnatelný s konvenčními metodami. Trendy výkonu metod se mezi datasety často výrazně liší.

Třetí test z článku [20] srovnává vybrané metody pomocí matching score, počtu nalezených korespondencí a počtu správných korespondencí. Dále byly zjištěny průměrné doby detekce a deskripce pro jeden významný bod. Pro srovnání je použitý dataset Oxfordské univerzity [6].

tab. 5 Počet korespondencí, počet špatných korespondencí, RC a čas potřebný pro detekci a deskripci jednoho významného bodu [20].

Metoda	Počet korespondencí	Počet špatných korespondencí	RC / %	Čas / η s
SIFT	470	85	82	90,44
SURF	178	52	71	41,83
KAZE	212	37	83	191,24
AKAZE	231	33	86	60,93
BRIEF	742	163	78	3,94
BRISK	544	114	79	16,59

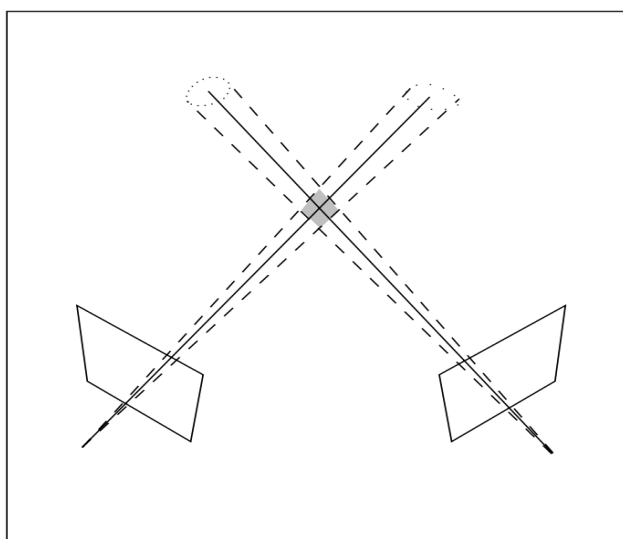
Z třetího srovnání má nejlepší hodnotu recall metoda AKAZE. Mezi jednotlivými metodami je velký rozdíl v počtu vytvořených korespondencí. Nevíce korespondencí bylo vytvořeno pomocí metody BRIEF, ale také touto metodou bylo detekováno nejvíce nesprávných korespondencí. Z časového srovnání je nejlepší metoda BRIEF, která několikanásobně překonává všechny ostatní metody. Naopak nejpomalejší metodou je metoda KAZE

3.1.7 Princip hledání korespondencí a triangulace

Jakmile jsou na snímcích detekovány a popsány významné body, musí se určit vzájemně odpovídající si páry bodů ze dvou snímků stejné scény, pořízených stereopárem. Pro každý významný bod z prvního snímku je nalezen významný bod z druhého snímku s nejvhodnějším popisem. K vytvoření korespondencí se využívá několika algoritmů, které se snaží co nejpresněji určit vhodné korespondující si body. Jeden z nejjednodušších způsobů, jak nalézt možné korespondující body, je povolit porovnávání jednoho významného bodu z prvního snímku se všemi významnými body druhého snímku. Korespondující body jsou učeny, pokud je popis významného bodu z levého snímku nejbližší popisu významného bodu ze všech bodů pravého snímku a stejně tak pro významný bod pravého snímku. Jiný způsob určení korespondujících si bodů je, že nalezneme dva nejbližší popisy a porovnáme je. Pokud je poměr rozdílu prvního a druhého popisu menší než určitá prahová hodnota, první popis je stabilní a významné body jsou určeny jako korespondující.

Dvojice bodů získané v předchozím kroku na základě podobnosti zpravidla obsahují mnoho nekorespondujících bodů (outliers), to nastává u bodů, které mají podobné popisy, ale neodpovídají jednomu fyzickému bodu ve scéně.

V případě, že jsou kamery kalibrované, nekorespondující si body jsou odstraněny na základě určení vzdálenosti mezi dvěma mimoběžnými přímkami. Přímký protínající středy kamer a korespondující si body v obraze jsou v důsledku nepřesností mimoběžky a zpravidla se neprotínou v jednom bodě, vzniká jistá oblast nejasností. Pro určení polohy bodu v prostoru je nalezen bod, který je těmto přímkám nejbližší na základě metody nejmenších čtverců. Pokud je vzdálenost mezi mimoběžkami větší než určitá prahová hodnota, nejedná o správně přiřazené body.



Obr. 3-26 Zobrazení oblasti nejasností [27].

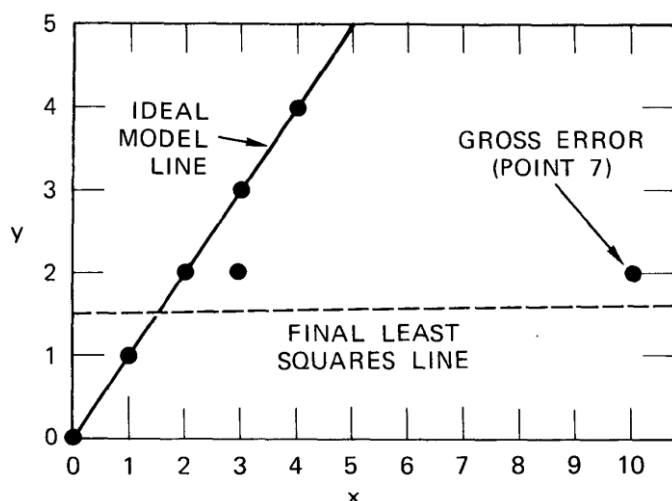
Kamery lze také kalibrovat na základě korespondencí následujícími algoritmy:

RANSAC (Random sample consensus) [28]. Tento algoritmus se náhodně snaží najít ideální transformaci mezi dvěma snímky. Pro nalezení správné projektivní transformace dvou rovin potřebujeme alespoň čtyři správně přiřazené významné body, které neleží na přímce. RANSAC vybírá pro výpočet odhadu transformace náhodně čtyři korespondující si body. Jako vstupní data se používají korespondence nalezené pomocí některé z výše uvedených metod. Je-li transformace spočtená ze správných korespondencí, měly by body jednoho snímku ležet na odpovídajících si místech druhého snímku. Kvalita transformace se hodnotí podle počtu korespondujících si bodů, které po transformaci leží v určité toleranci okolo místa, kde se ve druhém snímku nachází korespondující bod. Tento postup se iterativně opakuje a cílem je nalézt takové čtyři body, podle kterých se vytvoří transformace, která odpovídá co největšímu počtu bodů.

MLESAC (Maximum Likelihood Estimate Sample Consensus) [29]. Tato úprava nepoužívá konkrétní hodnotu tolerance, ale používá Gaussovu funkci a algoritmus se snaží najít takovou transformaci, při níž budou korespondující si body nejbližší.

PROSAC (Progressive Sample Consensus) [30]. Tato úprava se snaží minimalizovat výpočetní náročnost a s tím souvisí menší časová náročnost. Algoritmus je podobný jako RANSAC. Náhodně vybírá korespondující si body, aby z nich vytvořil transformaci. Na rozdíl od RANSACu nevybírá ze všech korespondujících si bodů, ale pouze z bodů, u kterých v první fázi byla zjištěna největší podobnost.

KALMANSAC [31]. Tato úprava využívá Kalmánův filtr a je schopna vytvořit transformaci dvou snímků jedné scény, na které se určité objekty pohybují.



Obr. 3-27 Zobrazení jednoduché úlohy, kdy 6 bodů leží téměř na jedné přímce a sedmý bod má velkou odchylku. Pomocí metody RANSAC je vytvořena ideální transformace (plná čára). Ve srovnání s metodou nejmenších čtverců (čárkovaná čára), kde výsledek není vyhovující [28].

3.2 Disparitní mapa

Algoritmy pro výpočet disparitní mapy nejčastěji porovnávají jas jednotlivých pixelů nebo skupin pixelů [32]. Jeden snímek je vždy označen za výchozí (base) a druhý za porovnávaný (match). Podle volby výchozího a porovnávaného snímku nazýváme metody left-to-right nebo right-to-left. Snímky musí být pořízeny z kalibrovaných kamer a musí být rektifikovány. Pro výpočet disparitní mapy se využívá mnoho algoritmů fungujících na podobném principu. Fáze výpočtu disparitní mapy lze rozdělit na rektifikaci, fázi výpočtu míry podobnosti a finální úpravu.

3.2.1 Rektifikace

Rektifikace je úprava snímku za účelem zjednodušení následného výpočtu míry podobnosti. Jedná se o transformaci obrazu z jednoho souřadného systému do souřadného systému druhého snímku. To znamená, že jednotlivé řady pixelů jednoho snímku odpovídají druhému snímku. Tímto se proces hledání korespondujících si bodů ve 2D

snímku zjednoduší na jednorozměrnou úlohu, protože se hledají pouze odpovídající si body na jednom řádku.

3.2.2 Výpočet míry podobnosti

V této fázi se vytváří vektor pro každý pixel nebo skupinu pixelů. Měří se rozdíl hodnoty jasu pixelu o souřadnicích $[x, y]$ ve výchozím snímku a hodnoty jasu pixelu o souřadnicích $[x + d, y]$ v porovnávaném snímku, kde “d” je velikost disparity testovaného pixelu. Výsledek této fáze představuje trojrozměrnou matici, kde dva rozměry představují souřadnice pixelů v obraze, a třetí rozměr znázorňuje disparitu vstupních obrazů. Z předpokladu, že se na obraze nachází větší plochy se stejnou hloubkou, respektive disparitou, metody porovnávají celé skupiny pixelů. Metody výpočtu míry podobnosti:

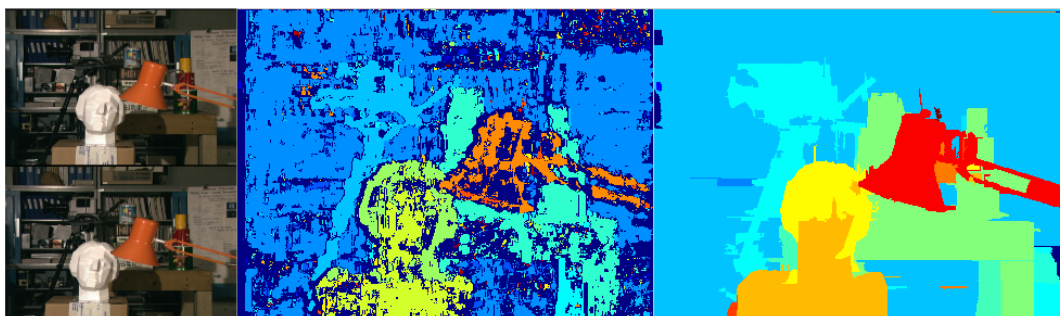
- Absolute intensity differences - absolutní hodnota rozdílů intenzit
- Squared intensity differences - suma kvadrátu rozdílů jasu
- Mutual Information - vzájemná informace

3.2.3 Finální úprava

Vypočtené disparity obsahují velké množství chyb, proto je potřeba hloubkovou mapu upravit. První úprava je agregace, jejímž cílem je odstranit drobné výkyvy, způsobené šumem a jinými nepřesnostmi obrazu. Agregace tedy vytváří souvislejší plochy a vyrovnává jednopixelové chyby.

- průměrování disparit jednotlivých skupin pixelů
- omezení maximální disparity pro omezení nereálných hodnot

Následuje určení výsledné neoptimálnější disparity. Existují dva přístupy, které lze využít - lokální a globální. Lokální metody hledají disparity jen v určitém okolí daného bodu. Tyto metody spočívají v nalezení disparity, při které mají pixely nejvyšší míru podobnosti. Globální metody hledají optimální disparity po celém řádku. Zohledňují při hledání shod kromě vlastností daného pixelu i to, jaké shody byly přiřazeny k okolním pixelům. Tedy přiřazení korespondujících bodů k sobě ovlivňuje přiřazování ostatních bodů v jejich okolí.



Obr. 3-28 Zobrazení původních obrazů, disparity jednotlivých pixelů a optimální disparity [4].

4 DISKUZE

V rámci bakalářské práce jsou popsány a srovnány přístupy a metody 3D rekonstrukce scény pomocí pasivního stereopáru kamer. Podmínkou pro pasivní 3D rekonstrukci je, že kamery musí být kalibrované a snímaná scéna musí být dostatečně strukturovaná. Jednotlivé přístupy se velmi liší svojí přesností, rychlostí a tedy svým možným uplatněním.

Prvním rozebíraným přístupem je 3D rekonstrukce scény pomocí triangulace významných bodů. Nezbytnou součástí tohoto přístupu je nalezení, popsání a konečná korespondence dvou významných bodů v obraze, které odpovídají jednomu reálnému bodu scény. Triangulace se používá pro přesné určení polohy jednotlivých bodů v prostoru. Nevýhodou je, že takto vznikne nespojitá množina bodů v prostoru a pro získání modelu scény musí být použito dalšího softwaru, jako je například Agisoft PhotoScan nebo PhotoModeler.

Metody pro detekci významných bodů se liší ve tvaru hledaných geometrických primitiv. Tímto je ovlivněna jejich výkonnost na různých typech scény, kdy se získané obrazy liší svojí strukturou (často se opakujícími geometrickými primitivy), například rohy, hrany, stabilní oblasti nebo místa, která představují styk dvou rohů. Toto ovlivňuje počet nalezených bodů v různých scénách. Proto je třeba volit vhodný detektor dle charakteru scény. Metody pro detekci významných bodů se dále liší v jejich stabilitě. Detektory musí být odolné na deformace obrazu, jako je rotace, změna měřítka, osvětlení nebo úhel pohledu. Metoda KAZE dosahovala téměř u všech porovnání nejlepších výsledků z hlediska stability nalezených významných bodů. Následovala metoda SURF, která dosahovala srovnatelných výsledků a ve specifickém případě byla dokonce lepší než KAZE. Naopak výrazně horších výsledků bylo dosaženo pomocí CNN metod TILDE a LIFT, které je navíc nutné trénovat. Proto se v současné době obecně používají spíše konvenční metody detekce významných bodů. Další důležitou vlastností je výpočetní čas, který může limitovat některé metody v možnosti jejich využití např. pro úkoly probíhající v reálném čase. V rámci snahy o optimalizaci a dosažení rychlejších výpočetních časů byla vytvořena např. konvoluční neuronová síť KCNN. Tato metoda je trénovaná oproti konvenčním metodám, které napodobuje, a po tréninku je s nimi srovnatelná. Při správné optimalizaci na specifickém hardwaru dosahuje řádově rychlejších výpočetních časů detekce než konvenční metody, které napodobuje. Díky KCNN je například možné použít metodu KAZE pro aplikace v reálném čase.

Metody pro popis významných bodů se liší v algoritmu popisu okolí významného bodu tak, aby byla možná následná korespondence dvou vzájemně odpovídajících si bodů. Požadavkem na tyto metody je co největší invariance vůči natočení, zkosení, změně měřítka, šumu, rozmazání či změně osvětlení. Ze srovnání jednotlivých metod pro deskripci významných bodů vyplývá, že neexistuje ideální metoda vhodná pro jakoukoliv aplikaci a výkon metody záleží na konkrétním obraze nebo typu deformací, viz rozdílné trendy výkonu jednotlivých metod na různých datasetech. Největší spolehlivosti často

dosahuje konvoluční neuronová síť LIFT, nevýhodou je nutnost trénování této sítě. Dále LIFT na datasetech, které se výrazně liší od těch, na kterých byl trénován, dosahuje nižšího výkonu, srovnatelného s konvenčními metodami. Alternativou může být metoda KAZE, která také dosahuje obecně výborné stability. Nevýhodou metody KAZE je výpočetní čas, který byl nejvyšší ze všech testovaných metod. Nejrychlejší metody jsou BRIEF, BRISK a FREAK, které využívají jednodušší formát popisu. Jejich jednoduchost se ale projevila snížením kvality popisu a výskytem většího množství špatně korespondovaných bodů.

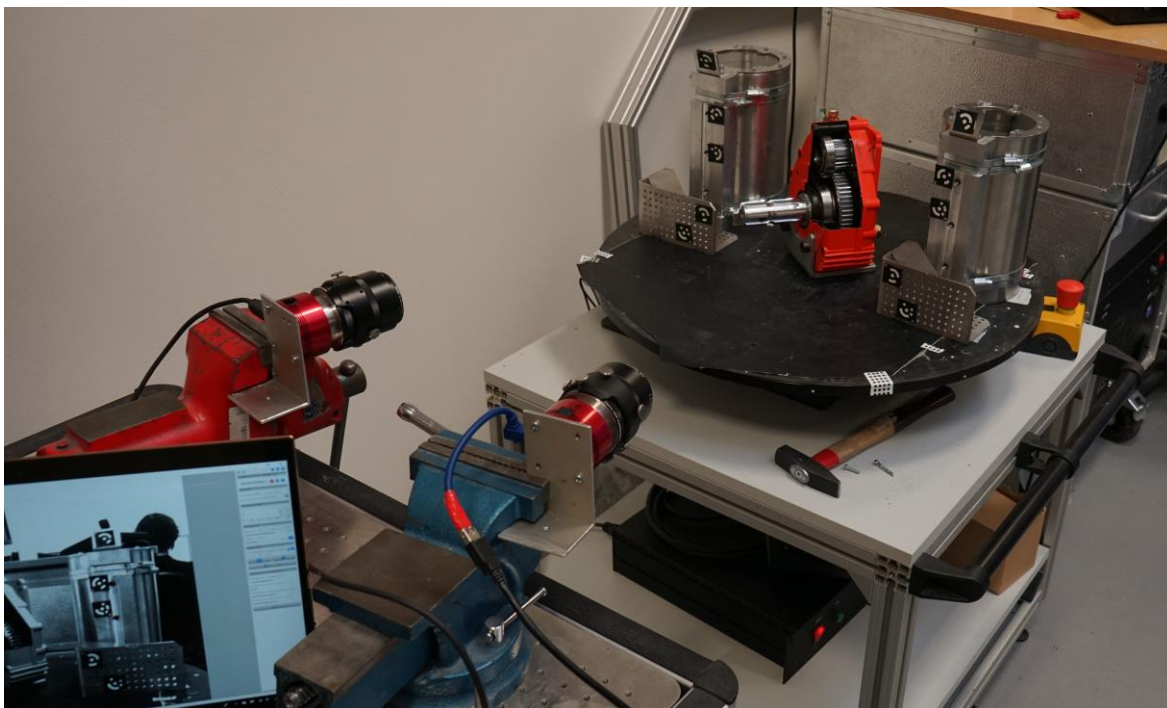
Neexistuje tedy univerzálně nejlepší metoda detekce ani deskripce a je nutné volit optimální metody v závislosti na konkrétní aplikaci. I přes rozvoj neuronových sítí se v praxi stále používají konvenční metody pro detekci i popis významných bodů, protože jsou s CNN stále srovnatelné a jejich použití je jednodušší.

Korespondence odpovídajících si bodů se vytvoří na základě nejpodobnějších popisů významných bodů. Tímto přiřazením však vznikne velké množství bodů, které neodpovídají jednomu bodu v prostotu a je nutné je odstranit. V případě, že známe parametry kamer, je možné tyto body nalézt pomocí přímek, které protínají středy kamer a korespondující si body v obraze. Omezením maximální vzdálenosti mezi těmito mimoběžkami lze odstranit nereálné hodnoty. V případě, že neznáme parametry kamer, je pomocí algoritmu RANSAC možné určit správnou transformaci mezi oběma obrazy, a tím stereopár kalibrovat.

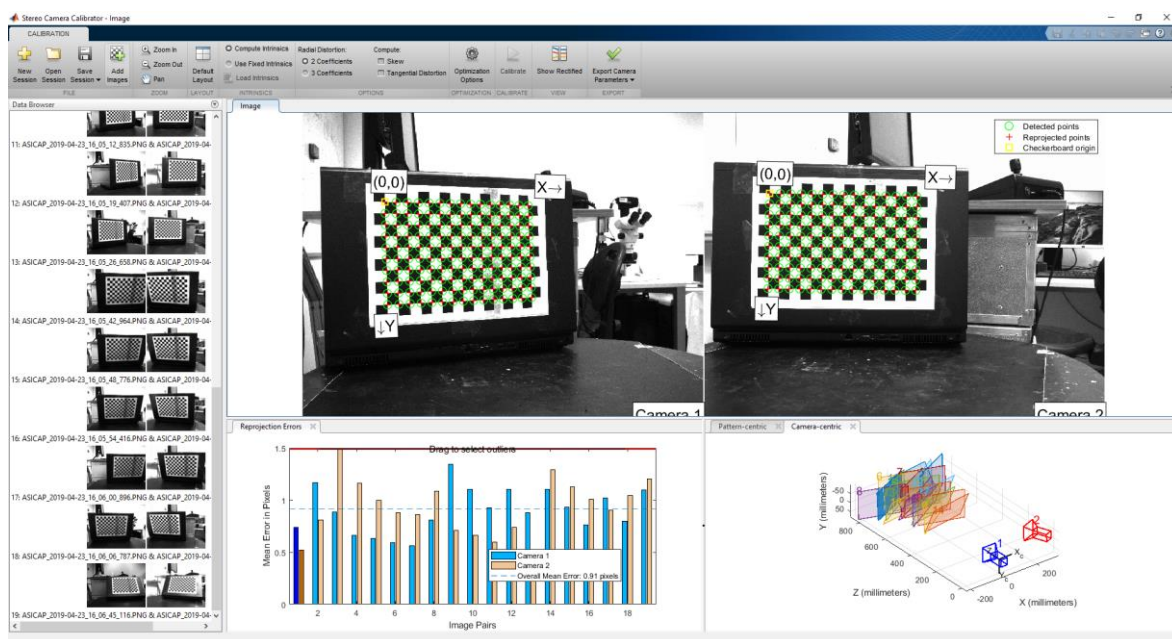
Druhým rozebíraným přístupem je disparitní mapa. Algoritmy pro výpočet disparitní mapy hledají nejpodobnější pixely nebo skupiny pixelů nejčastěji na základě rozdílu jejich jasu. Nezbytnou součástí tohoto přístupu je rektifikace, díky které jednotlivé řady pixelů jednoho snímku odpovídají řadám druhého snímku. Následuje vlastní výpočet disparity, kdy algoritmus porovnává hodnoty jasu pro jednotlivá posunutí obrazu v určitém rozsahu. Výsledek si lze představit jako trojrozměrnou matici, kde dva rozměry představují souřadnice pixelů, a třetí rozměr znázorňuje jejich posunutí. Tento výsledek je upraven a je určeno optimální posunutí, které má s porovnávaným pixelem největší podobnost. Velikost posunutí lze znázornit barevnou škálou a tím získáme vjem hloubky scény. Výpočet disparitní mapy je rychlý a může tak sloužit k navigaci robotů autonomních automobilů nebo pro aplikace jako je AR nebo VR. Nevýhodou je, že není schopný vypočítat přesnou polohu bodu v prostoru.

4.1 Praktický příklad

Pro praktický příklad byly použity snímky zachycené pomocí kamer ZWO ASI1600MM. Kamery byly uchycené pomocí svěráků, aby se zamezilo jejich pohybu. Stereopár byl zkalibrován pomocí kalibrační šachovnice. Pro výpočty byl použit software MATLAB s rozšířením Computer Vision System Toolbox.

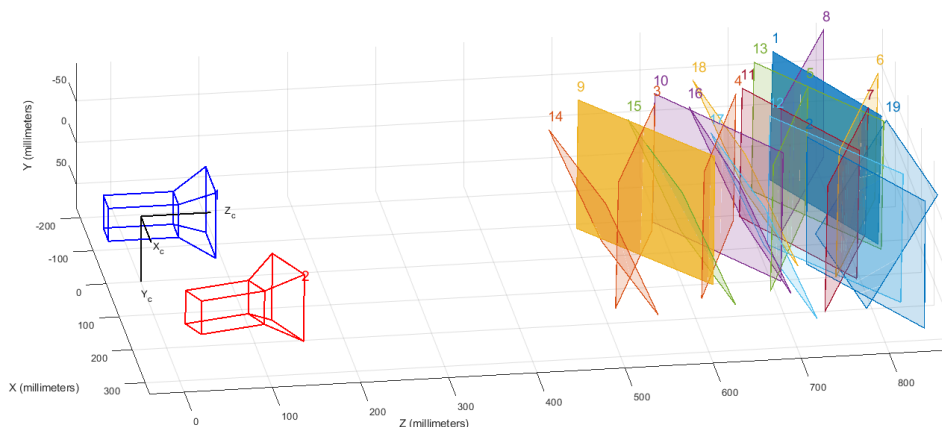


Obr. 4-1 Rozmístění aparatury a snímaná scéna.



Obr. 4-2 Grafické prostředí aplikace stereo camera calibrator.

Ke kalibraci bylo použito celkem 19 párů různě pootočených a posunutých snímků, dle doporučení aplikace tak, aby byl pokryt celý měřicí objem. Střední hodnota odchylky byla 0,91 pixelu, což je relativně vysoká hodnota, která byla způsobena tím, že kalibrační šachovnice byla vytisknuta běžnou stolní tiskárnou. Pro účely ilustračního příkladu je tato hodnota dostačující.



Obr. 4-3 Zobrazení kamer a polohy kalibrační šachovnice na jednotlivých snímcích.

Následně byl vytvořen kód pro triangulaci významných bodů.

```

1 - I1 = imread('img1.png');
2 - I2 = imread('img2.png');
3
4 - I1 = imgaussfilt(undistortImage(I1, stereoParams.CameraParameters1), 1);
5 - I2 = imgaussfilt(undistortImage(I2, stereoParams.CameraParameters2), 1);
6
7 - points1 = detectSURFFeatures(I1);
8 - points2 = detectSURFFeatures(I2);
9 - [features1, vpts1] = extractFeatures(I1, points1);
10 - [features2, vpts2] = extractFeatures(I2, points2);
11
12 - indexPairs = matchFeatures(features1, features2);
13 - matchedPoints1 = vpts1(indexPairs(:, 1));
14 - matchedPoints2 = vpts2(indexPairs(:, 2));
15
16 - [worldPoints, reproErrors] = triangulate(matchedPoints1, matchedPoints2, stereoParams);
17 - figure; showMatchedFeatures(I1, I2, matchedPoints1, matchedPoints2);
18
19 - T = 900;
20 - ii=1;
21 - for i = 1:size(reproErrors, 1)
22 -     if reproErrors(ii, 1) > 5 || worldPoints(ii, 3) > T || worldPoints(ii, 2) > T || ...
23 -         worldPoints(ii, 1) > T || worldPoints(ii, 3) < -T || worldPoints(ii, 2) < -T || worldPoints(ii, 1) < -T
24 -         worldPoints(ii, :) = [];
25 -         reproErrors(ii, :) = [];
26 -         matchedPoints1(ii, :) = [];
27 -         matchedPoints2(ii, :) = [];
28 -     else
29 -         ii=ii+1;
30 -     end
31 - end
32
33 - imshow(I1); hold on;
34 - plot(matchedPoints1);
35
36 - figure; showMatchedFeatures(I1, I2, matchedPoints1, matchedPoints2);
37
38 - figure; ptCloud = pointCloud(worldPoints);
39 - pcshow(ptCloud, 'MarkerSize', 80);
40

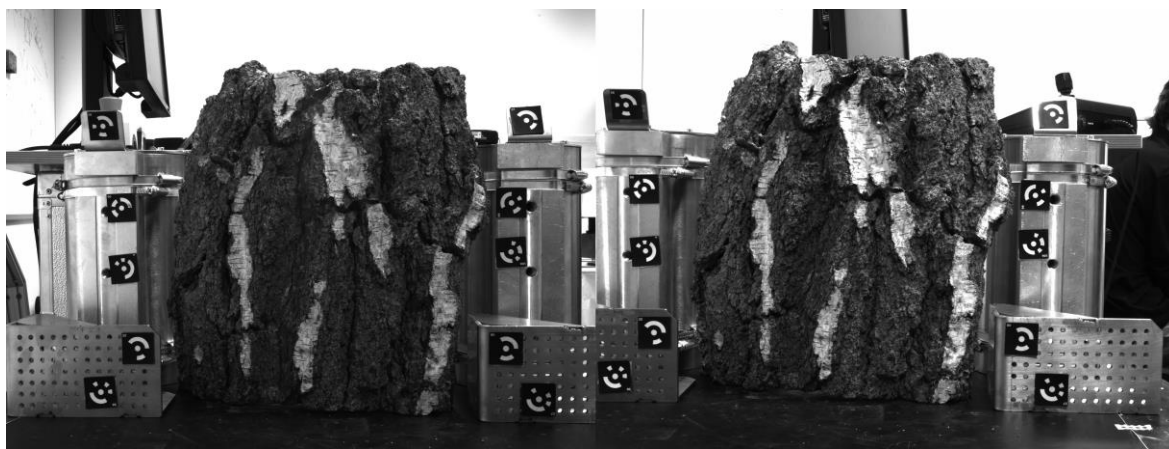
```

Obr. 4-4 Zobrazení kódu v editoru.

Kód obsahuje následující části (chronologicky):

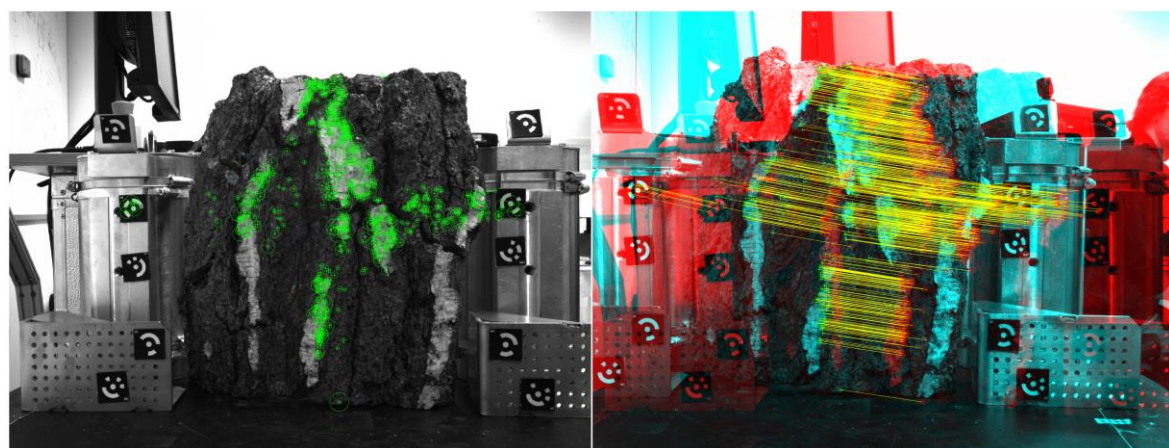
- Načtení snímků z obou kamer.
- Odstranění zkreslení objektivů a odstranění šumu pomocí Gaussova filtru.
- Detekce bodů pomocí běžně používaného detektoru SURF a vytvoření popisů.
- Přiřazení korespondujících si významných bodů.
- Triangulace korespondujících si bodů.
- Odstranění špatně přiřazených bodů a bodů, ležících daleko od snímaného objektu.
- Vykreslení korespondujících si bodů na anaglyfu.
- Vykreslení bodů v prostoru.

Jako snímaný objekt byla vybrána kůra (borka), kvůli dostatečně strukturovanému povrchu. Na **Obr. 3-1** je pár použitých snímků.

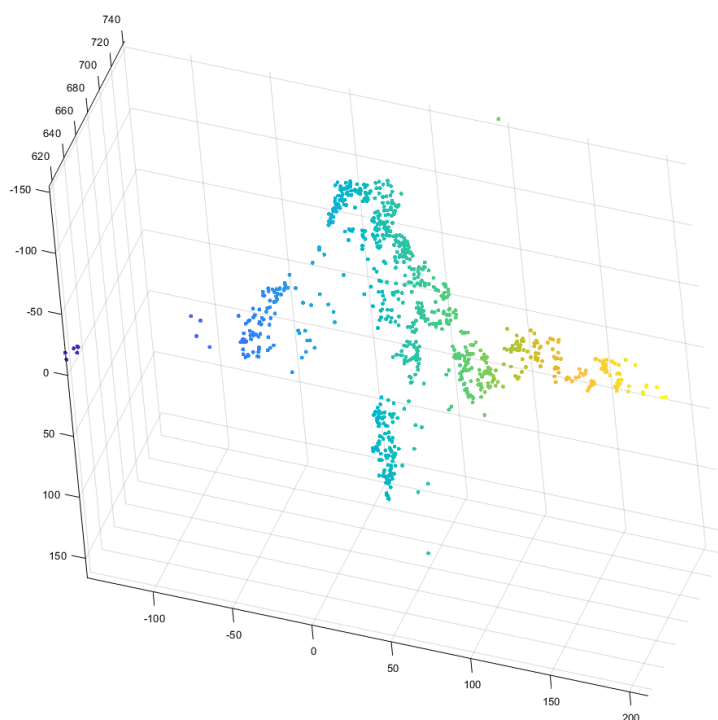


Obr. 4-5 Zobrazení originálních snímků.

Ve více strukturovaných částech obrázku byla nalezena hustá síť významných bodů, jak je možné vidět na **Obr. 4-6**.



Obr. 4-6 Zobrazení nalezených významných bodů na jednom snímku (vlevo) a korespondujících si bodů na anaglyfu (vpravo).



Obr. 4-7 Zobrazení jednotlivých bodů v prostoru.

Na **Obr. 4-7** je vidět výsledek praktického příkladu 3D rekonstrukce, při němž byly zjištěny souřadnice jednotlivých významných bodů v prostoru. Díky strukturovanému povrchu objektu byla nalezena hustá síť významných bodů. Pomocí případné úpravy některých parametrů detektoru by bylo možné nalézt rovnoměrněji rozložené významné body a ještě lépe zachytit scénu. Takto vypočítané souřadnice bodů mohou být exportovány a následně využity pro konkrétní účely.

5 ZÁVĚR

V bakalářské práci byl vytvořen aktuální přehled a srovnání jednotlivých přístupů a metod pro zpracování obrazových dat z pasivního stereopáru za účelem 3D rekonstrukce scény. Tyto přístupy a metody byly srovnány z hlediska jejich funkce, spolehlivosti, rychlosti a použitelnosti. Disparitní mapa je vhodná pro odhad hloubky scény v reálném čase, zatímco triangulace významných bodů přesně určuje polohu významných bodů v prostoru. Dále byly srovnány jednotlivé metody v rámci přístupu triangulace významných bodů. Na základě srovnání a znalosti principů těchto algoritmů byly vyvozeny informace o vhodnosti volby jednotlivých metod pro různé aplikace. Mezi jednotlivými metodami byly násobné rozdíly ve všech klíčových vlastnostech. Nelze proto určit univerzálně nejvhodnější metodu. Je třeba zvážit specifika konkrétní aplikace, včetně požadavků na rychlost, typ geometrických primitiv zachycených na obrázku nebo požadovanou odolnost vůči změně obrazu. Byl také vytvořen praktický příklad ilustrující funkci těchto přístupů.

Tato práce může čtenáři přinést vhled do problematiky zpracování obrazových dat z pasivního stereopáru za účelem 3D rekonstrukce scény. Na základě informací shrnutých v této práci může být čtenářem zvolen vhodný přístup nebo vhodná metoda s přihlédnutím ke specifickým vlastnostem konkrétní aplikace.

Rekonstrukce 3D scény pomocí stereopáru kamer má v dnešní době mnoho využití v širokém spektru odvětví. Aplikace jejich metod vede v průmyslu ke zvýšení efektivity výroby díky automatizování některých procesů. Dále může být použita pro navigaci autonomních automobilů a robotů, tvorbu modelů pro 3D tisk, pro herní a filmový průmysl. A v neposlední řadě pro medicínské aplikace, kdy na základě 3D rekonstrukce může být pro pacienta zhotoven implantát na míru. S příchodem konvolučních neuronových sítí a zlepšováním kvality algoritmů počítačového vidění se rozšiřuje možnost uplatnění 3D rekonstrukce do dalších odvětví.

6 SEZNAM POUŽITÝCH ZDROJŮ

- [1] 3D stereo machine vision depth sensing. *Environmental Engineering News Online* [online]. Copyright © 2019 [cit. 13.05.2019]. Dostupné z: <https://www.environmentalengineering.org.uk/news/3d-stereo-machine-vision-depth-sensing-6099/>.
- [2] LUHMANN, Thomas, Stuart ROBSON, Stephen KYLE a Jan BOEHM. *Close Range Photogrammetry and 3D Imaging*. 2013. ISBN 978-3110302691.
- [3] FUHRMANN, Simon, Fabian LANGGUTH, Nils MOEHRLE, Michael WAECHTER a Michael GOESELE. MVE—An image-based reconstruction environment. *Computers & Graphics*. 2015, **53**, 44–53. ISSN 0097-8493. doi:<https://doi.org/10.1016/j.cag.2015.09.003>
- [4] *Sucuri WebSite Firewall - Access Denied*. *Sucuri WebSite Firewall - Access Denied* [online]. Copyright © 2019 Sucuri Inc. All rights reserved. [cit. 13.05.2019]. Dostupné z: <http://www.shawnlankton.com/2007/12/3d-vision-with-stereo-disparity/>.
- [5] KAHAKI, Seyed Mostafa Mousavi, Md Jan NORDIN a Amir Hossein ASHTARI. Contour-Based Corner Detection and Classification by Using Mean Projection Transform. *Sensors*. 2014, **14**(3), 4126–4143. ISSN 1424-8220. doi:[10.3390/s140304126](https://doi.org/10.3390/s140304126)
- [6] MIKOLAJCZYK, Krystian, Tinne TUYTELAARS, Cordelia SCHMID, Andrew ZISSERMAN, Jiri MATAS, Frederik SCHAFFALITZKY, Timor KADIR a Luc VAN GOOL. A Comparison of Affine Region Detectors. *International Journal of Computer Vision*. 2005, **65**, 43–72. doi:[10.1007/s11263-005-3848-x](https://doi.org/10.1007/s11263-005-3848-x)
- [7] BEDROS, Saad J. Lecture 8: Interest Point Detection [online]. 2015. Dostupné z: <http://me.umn.edu/courses/me5286/vision/Notes/2015/ME5286-Lecture8.pdf>
- [8] RODEHORST, V a Andreas KOSCHAN. Comparison and evaluation of feature point detectors. 2019.
- [9] SMITH, S M a J M BRADY. SUSAN - A new approach to low level image processing. *International Journal of Computer Vision*. 1997, **23**(1), 45–78. ISSN 0920-5691. doi:[10.1023/a:1007963824710](https://doi.org/10.1023/a:1007963824710)
- [10] LOWE, D G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*. 2004, **60**(2), 91–110. ISSN 0920-5691. doi:[10.1023/b:visi.0000029664.99615.94](https://doi.org/10.1023/b:visi.0000029664.99615.94)
- [11] BAY, H, A ESS, T TUYTELAARS a L VAN GOOL. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*. 2008, **110**(3), 346–359. ISSN 1077-3142. doi:[10.1016/j.cviu.2007.09.014](https://doi.org/10.1016/j.cviu.2007.09.014)
- [12] ALCANTARILLA, P F, A BARTOLI a A J DAVISON. KAZE Features. *Computer Vision - Eccv 2012, Pt Vi*. 2012, **7577**, 214–227. ISSN 0302-9743.
- [13] ROSTEN, E, T DRUMMOND a S O C IEEE COMPUTER. Fusing points and lines for high performance tracking. In: *10th IEEE International Conference on Computer Vision (ICCV 2005)*. LOS ALAMITOS: Ieee Computer Soc, 2005, s. 1508–1515. IEEE International Conference on Computer Vision. ISBN 0-7695-2334-X.

- [14] MATAS, J, O CHUM, M URBAN a T PAJDLA. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*. 2004, **22**(10), 761–767. ISSN 0262-8856. doi:10.1016/j.imavis.2004.02.006
- [15] BIMBO, Alberto Del. Region detectors Requirements for region detection. *MSER and Region detector* [online]. 2011, 10. Dostupné z: http://www.micc.unifi.it/delbimbo/wp-content/uploads/2011/03/slide_corso/A34_MSER.pdf
- [16] ALDANA-IUIT, J, D MISHKIN, O CHUM, J MATAS a IEEE. In the Saddle: Chasing Fast and Repeatable Features. In: *23rd International Conference on Pattern Recognition (ICPR)*. LOS ALAMITOS: Ieee Computer Soc, 2016, s. 675–680. International Conference on Pattern Recognition. ISBN 978-1-5090-4847-2.
- [17] VERDIE, Y, K M YI, P FUA, V LEPETIT a IEEE. TILDE: A Temporally Invariant Learned DEtector. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. NEW YORK: Ieee, 2015, s. 5279–5288. IEEE Conference on Computer Vision and Pattern Recognition. ISBN 978-1-4673-6964-0.
- [18] YI, K M, E TRULLS, V LEPETIT a P FUA. LIFT: Learned Invariant Feature Transform. *Computer Vision - Eccv 2016, Pt Vi*. 2016, **9910**, 467–483. ISSN 0302-9743. doi:10.1007/978-3-319-46466-4_28
- [19] DI FEBBO, P, C DAL MUTTO, K TIEU, S MATTOCCIA a IEEE. KCNN: Extremely-Efficient Hardware Keypoint Detection with a Compact Convolutional Neural Network. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. NEW YORK: Ieee, 2018, s. 795–803. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. ISBN 978-1-5386-6100-0. doi:10.1109/cvprw.2018.00111
- [20] TAREEN, S A K a Z SALEEM. A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK. In: *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*. 2018, s. 1–10. doi:10.1109/ICOMET.2018.8346440
- [21] ALCANTARILLA, P F, J NUEVO a A BARTOLI. Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces. *Proceedings of the British Machine Vision Conference 2013*. 2013, 11. doi:10.5244/c.27.13
- [22] CALONDER, M, V LEPETIT, C STRECHA a P FUA. BRIEF: Binary Robust Independent Elementary Features. *Computer Vision-Eccv 2010, Pt Iv*. 2010, **6314**, 778–792. ISSN 0302-9743.
- [23] LEUTENEGGER, S, M CHLI, R Y SIEGWART a IEEE. BRISK: Binary Robust Invariant Scalable Keypoints. In: *IEEE International Conference on Computer Vision (ICCV)*. NEW YORK: Ieee, 2011, s. 2548–2555. IEEE International Conference on Computer Vision. ISBN 978-1-4577-1102-2.
- [24] ALAHI, A, R ORTIZ, P VANDERGHEYNST a IEEE. FREAK: Fast Retina Keypoint. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. NEW YORK: Ieee, 2012, s. 510–517. IEEE Conference on Computer Vision and Pattern Recognition. ISBN 978-1-4673-1228-8.

- [25] STRECHA, Christoph, Wolfgang VON HANSEN, Luc VAN GOOL, Pascal FUA a Ulrich THOENNESSEN. On Benchmarking Camera Calibration and Multi-View Stereo for High Resolution Imagery. *CVPR*. 2008. doi:10.1109/CVPR.2008.4587706
- [26] AANÆS, Henrik, Anders Lindbjerg DAHL a Kim STEENSTRUP PEDERSEN. Interesting Interest Points. *Int. J. Comput. Vision*. 2012, **97**(1), 18–35. ISSN 0920-5691. doi:10.1007/s11263-011-0473-8
- [27] HARTLEY, Richard a Andrew ZISSERMAN. *Multiple View Geometry in Computer Vision*. 2. vyd. New York, NY, USA: Cambridge University Press, 2003. ISBN 0521540518.
- [28] FISCHLER, M A a R C BOLLES. RANDOM SAMPLE CONSENSUS - A PARADIGM FOR MODEL-FITTING WITH APPLICATIONS TO IMAGE-ANALYSIS AND AUTOMATED CARTOGRAPHY. *Communications of the Acm*. 1981, **24**(6), 381–395. ISSN 0001-0782. doi:10.1145/358669.358692
- [29] TORR, P H S a A ZISSERMAN. MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*. 2000, **78**(1), 138–156. ISSN 1077-3142. doi:10.1006/cviu.1999.0832
- [30] CHUM, O a J MATAS. Matching with PROSAC - Progressive Sample Consensus. In: *Conference on Computer Vision and Pattern Recognition*. LOS ALAMITOS: Ieee Computer Soc, 2005, s. 220–226. IEEE Conference on Computer Vision and Pattern Recognition. ISBN 0-7695-2372-2.
- [31] VEDALDI, A, H L JIN, P FAVARO, S SOATTO a S O C IEEE COMPUTER. KALMANSAC: Robust filtering by consensus. In: *10th IEEE International Conference on Computer Vision (ICCV 2005)*. LOS ALAMITOS: Ieee Computer Soc, 2005, s. 633–640. IEEE International Conference on Computer Vision. ISBN 0-7695-2334-X.
- [32] TRUCCO, Emanuele, K PLAKAS, Nicole BRANDENBURG, Peter KAUFF, Michael KARL a Oliver SCHREER. Real-Time Disparity Maps for Immersive 3-D Teleconferencing by Hybrid Recursive Matching and Census Transform. 2001.

7 SEZNAM POUŽITÝCH ZKRATEK

CNN	Konvoluční neuronová síť
MS	Matching score - poměr počtu správných korespondencí vůči počtu nalezených významných bodů
RC	Recall - poměr počtu správných korespondencí vůči celkovému počtu korespondencí
VR	Virtuální realita
AR	Rozšířená realita

8 SEZNAM OBRÁZKŮ

Obr. 3-1 Pasivní stereopár kamer [1].	15
Obr. 3-2 Zobrazení parametrů kamery [2].	16
Obr. 3-3 Významné body (vlevo), korespondující si body ze dvou snímků (vpravo) [3].	17
Obr. 3-4 Disparitní mapa a originální obrazy scény [4].	17
Obr. 3-5 Významné body nalezené detektorem rohů [5].	18
Obr. 3-6 Zobrazení náchylnosti Moravcova detektoru na rotaci [7].	19
Obr. 3-7 Náchylnost Harrisova detektoru na změnu měřítka [7].	20
Obr. 3-8 Pět kruhových oblastí s vyznačeným středem různě umístěných na jednoduchém obraze (vlevo). Pět stejných kruhových oblastí. Znázornění oblasti USAN bílou barvou (vpravo) [9].	20
Obr. 3-9 Trojrozměrný graf zobrazující velikosti oblasti USAN v závislosti na pozici vyšetřovaného bodu na obraze [9].	21
Obr. 3-10 Tvorba scale space a výpočet DoG [10].	21
Obr. 3-11 Druhé derivace Gaussovy funkce podle y a podle xy s použitím čtvercových filtrů. (šedá pole mají hodnotu 0) [11].	22
Obr. 3-12 Porovnání Gaussova filtru (nahore) a nelineárního difuzního filtru (dole) pro různá zkruslení [12].	22
Obr. 3-13 Bresenhamova kružnice s 16 vyznačenými pixely a středem (c), body na pozicích 1,9,5 a 13 jsou vyšetřovány jako první [13].	23
Obr. 3-14 Znázornění prahování obrazu [15].	24
Obr. 3-15 Zobrazení oblastí nalezených pomocí MSER detektoru [14].	24
Obr. 3-16 Znázornění Bresenhamovy kružnice (modré pixely) a vnitřní kružnice (růžové body) (vlevo). Čtyři přípustné kombinace jasu pixelů z první fáze detekce (uprostřed). Příklady detektorem vyhodnocených významných bodů (vpravo). Barvy odpovídají skupinám: světlejší oblast (l) - zelená, tmavší oblast (d) - červená, oblast podobného jasu (s) - modrá [16].	25
Obr. 3-17 Architektura sítě KCNN. I je vstupní obrázek a $p(I)$ je výsledná odezva významných bodů. Každé okno znázorňuje konvoluční operaci. Filtry v první konvoluční vrstvě jsou oddělitelné [19].	27
Obr. 3-18 Znázornění kompaktního hardware systému, na kterém je implementován KCNN [19].	27

Obr. 3-19 Zobrazení invariance jednotlivých metod na rozmazání, změnu úhlu pohledu, měřítko s rotací a osvětlení. Vertikální osa znázorňuje míru opětovného nalezení bodu. Horizontální osa znázorňuje míru zkreslení [19].	28
Obr. 3-20 Zjednodušená ukázka deskriptoru SIFT [10].	30
Obr. 3-21 Haarova vlnka (vlevo) zobrazení dominantní orientace (vpravo) [11].	31
Obr. 3-22 Odezva Haarovy vlny na různé vzory [11]	32
Obr. 3-23 Náhodně propojené dvojice bodů, symetricky propojené body (vpravo dole) [22].	33
Obr. 3-24 Znázornění soustředných kružnic ležících okolo významného bodu [23].	33
Obr. 3-25 Rozložení porovnávaných bodů použitých k určení orientace deskriptoru [24].	34
Obr. 3-26 Zobrazení oblasti nejasností [27].	38
Obr. 3-27 Zobrazení jednoduché úlohy, kdy 6 bodů leží téměř na jedné přímce a sedmý bod má velkou odchylku. Pomocí metody RANSAC je vytvořena ideální transformace (plná čára). Ve srovnání s metodou nejmenších čtverců (čárkovaná čára), kde výsledek není vyhovující [28].	39
Obr. 3-28 Zobrazení původních obrazů, disparity jednotlivých pixelů a optimální disparity [4].	40
Obr. 4-1 Rozmístění aparatury a snímaná scéna.	43
Obr. 4-2 Grafické prostředí aplikace stereo camera calibrator.	43
Obr. 4-3 Zobrazení kamer a polohy kalibrační šachovnice na jednotlivých snímcích.	44
Obr. 4-4 Zobrazení kódu v editoru.	44
Obr. 4-5 Zobrazení originálních snímků.	45
Obr. 4-6 Zobrazení nalezených významných bodů na jednom snímku (vlevo) a korespondujících si bodů na anaglyfu (vpravo).	45
Obr. 4-7 Zobrazení jednotlivých bodů v prostoru.	46

9 SEZNAM TABULEK

tab. 1 Zobrazení doby detekce významných bodů pro metody KAZE a KCNN na různých procesorech [19].	29
tab. 2 Srovnání počtu významných bodů nalezených vybranými detektory (průměr ze dvou datasetů) [20].	29
tab. 3 Srovnání deskriptorů podle MS a RC [21].	36
tab. 4 Průměrné Matching score pro vybrané deskriptory [18].	36
tab. 5 Počet korespondencí, počet špatných korespondencí, RC a čas potřebný pro detekci a deskripci jednoho významného bodu [20].	37